Nonlinear Taxation and International Mobility in General Equilibrium*

Eckhard Janeba† **Karl Schulz**‡

University of Mannheim University of St. Gallen

November 24, 2022

Abstract

Migration and general equilibrium forces are both known to limit the extent of redistribution due to a migration threat and a trickle-down rationale, respectively. In this paper, we consider these two forces jointly and study the optimal nonlinear taxation of internationally mobile workers in general equilibrium. We show that both forces partly offset each other. In general equilibrium, migration may lower the bottom tax rate but raises the top tax rate, challenging the classical migration-threat argument. Moreover, we demonstrate that migration responses weaken the trickle-down rationale. Both findings can be explained by a novel wage effect on migration and a migration effect on wages, calling for higher top tax rates to amplify pre-tax wage inequality and prevent high-skilled emigration. We calibrate our model to the U.S. economy and illustrate the new effects by comparing the optimal tax schemes with and without migration, as well as with and without endogeneity of wages.

Keywords: Optimal Taxation, General Equilibrium, Trickle-Down Effects, Migration, Tax/Subsidy Competition

JEL Classification: H21, H24, H73, F22, R13

*Eckhard Janeba gratefully acknowledges the support from the Collaborative Research Center (SFB) 884 "Political Economy of Reforms", funded by the German Research Foundation (DFG). For helpful comments and discussions, we thank Hans Peter Grüner, Duk Gyoo Kim, Mathilde Muñoz, Dominik Sachs, Sebastian Siegloch, Nicolas Werquin, and numerous seminar and conference participants.

†University of Mannheim, Department of Economics, L7 3-5, 68131 Mannheim, Germany, janeba@uni-mannheim.de.

‡University of St. Gallen, Department of Economics, Bodanstrasse 8, 9000 St. Gallen, Switzerland, karl.schulz@unisg.ch.

1 Introduction

International (and inter-regional) mobility of high-income individuals has been at the center of recent theoretical and empirical research due to its far-reaching implications for the taxation of mobile individuals and the progressivity of the income tax code. Mirrlees (1971) has already recognized the importance of international migration but focuses on a closed economy case in his formal analysis.[1] The more recent literature on optimal income taxation has demonstrated the effects of migration on the level and shape of optimal marginal income tax rates. In particular, migration tends to decrease optimal marginal tax rates over the entire income distribution (see Simula and Trannoy (2010)) and, depending on the shape of migration semi-elasticities, leads to negative tax rates at the top (Lehmann, Simula, and Trannoy (2014)) and reduces tax payments by top-income workers relative to a closed economy. The driving force is a "threat of migration" that is even present in a situation in which no net migration occurs in equilibrium. A standard assumption in these models is that workers' wages are exogenous. However, migration responses are known to affect the distribution of wages substantially (e.g., Dustmann, Frattini, and Preston (2013)). This should be considered when evaluating the impact of labor mobility on the progressivity of income tax schedules.

The role of wage endogeneity for tax policy has been highlighted in another strand of the optimal income taxation literature. For instance, in the closed economy model of Stiglitz (1982), the government lowers the top tax rate to encourage the labor supply of high-skilled individuals, thereby raising wages at the bottom and reducing them at the top—a "trickle-down" effect.[2] Contributors to the literature refer to this indirect redistribution of pre-tax wages as "predistribution." However, this literature has so far disregarded the empirically relevant role of tax-induced labor mobility, in particular of high-income earners (discussed below).

This paper bridges the gap between these two strands of the literature by studying optimal income taxation and international mobility in general equilibrium, i.e., with endogenous wages. Our main result is that migration and general equilibrium effects partly offset each other. In other words, endogenous wages limit the incentive to cut taxes on mobile, high-income earners, and the presence of migration weakens predistribution. We develop this finding in two steps. As

[1] See Kleven, Landais, Munoz, and Stantcheva (2020) for a survey of the empirical literature.

[2] While Stiglitz (1982) makes this observation about general equilibrium wage effects in a two-type framework, Sachs, Tsyvinski, and Werquin (2020) demonstrate in a continuous-type framework that lower tax rates in general equilibrium may also apply to middle-class workers (see Figure 4 in Sachs et al. (2020)).

Migration Threat: top tax rate $\downarrow \rightsquigarrow$ top emigration $\downarrow \rightsquigarrow$ tax base $\uparrow$
Trickle Down: top tax rate $\downarrow \rightsquigarrow$ top labor supply $\uparrow \rightsquigarrow$ pre-tax wage inequality $\downarrow$
Wage Effect on Migration: top tax rate $\uparrow \rightsquigarrow$ pre-tax wage inequality $\uparrow \rightsquigarrow$ top immigration, tax base $\uparrow$
Migration Effect on Wages: top tax rate $\uparrow \rightsquigarrow$ top emigration $\uparrow \rightsquigarrow$ pre-tax wage inequality, tax base $\uparrow$

Table 1: Mechanisms: Taxes, Migration, and Endogenous Wages

a first step, we introduce labor mobility (as in Lehmann et al. (2014)) into a generalized version of the Stiglitz (1982) model of income taxation in a closed economy, here with K productivity types.[3] We show that in general equilibrium, the optimal marginal tax rate at the top is higher (i.e., a lower labor subsidy) in the open than in the closed economy. Moreover, we derive conditions under which the optimal marginal tax rate at the bottom declines. Thus, migration may lead to a more progressive tax code in terms of marginal tax rates.

Then, we turn to a continuous-type framework (e.g., Mirrlees (1971)) and derive the non-linear tax incidence and optimal taxation in general equilibrium with labor mobility. While we recover the standard migration threat (see Lehmann et al. (2014)) and trickle-down effect (see Sachs et al. (2020)) limiting the extent of optimal redistribution, we discover two novel effects that capture the interaction of migration and general equilibrium effects: a wage effect on migration and a migration effect on wages. In Table 1, we summarize the main mechanisms.

Table 1 allows us to describe the novel effects in our model resulting from a tax increase. On the one hand, a higher tax on, say, high-skilled workers, lowers their labor supply, thereby raising pre-tax wages at the top and lowering them at the bottom. The government predistributes less to low-skilled workers. This rise in pre-tax wage inequality triggers an immigration response of high-skilled workers and broadens the tax base. We label this immigration triggered by the higher wage inequality as a *wage effect on migration*. On the other hand, following a rise in the top tax rate, high-skilled workers migrate out a country and, thus, aggregate labor supply at the top declines. As a result, high-skilled pre-tax wages go up, while those at the

[3] Muting migration responses, the setup closely resembles the one in Ales, Kurnaz, and Sleet (2015).

bottom decline. This migration-induced increase in pre-tax wage inequality expands the tax base—a *migration effect on wages*. As a consequence, the migration-threat argument is less severe in general equilibrium, since any tax-induced out-migration of high-skilled workers amplifies wage inequality, thereby raising tax revenues. Moreover, the usual trickle-down effect, through which a tax cut at the top raises low-skilled wages, is self-limiting under labor mobility: Any tax-induced reduction in pre-tax wage inequality comes along with lower immigration (or higher emigration) of high-skilled workers, which, in turn, raises their wages, partly offsetting the benefits of trickle down.

We calibrate our continuous-type model to the U.S. economy and simulate the optimal tax schemes across different technologies (with vs. without migration, with vs. without endogenous wages). In line with our theoretical argument, the adjustment for migration is much smaller in general than in partial equilibrium. Migration responses may even lead to higher tax rates for the majority of workers. Similarly, with migration, the presence of general equilibrium responses only slightly changes the optimal tax scheme, whereas, without migration, it would lead to substantial tax rebates at the top. Altogether, both the migration threat and the trickle-down rationale appear exaggerated when accounting for the interaction of migration and general equilibrium responses.

Related literature. Our work is related to several important contributions to the literature on optimal nonlinear income taxation. Mirrlees (1971), Diamond (1998), and Saez (2001) study the optimal taxation of endogenous incomes without migration and with exogenous wages. Stiglitz (1982) initiated the debate on labor income taxation with endogenous wages in a two-type setting. As Ales et al. (2015), we study the optimal taxation in a K-type version of the Stiglitz (1982) model using a constant elasticity production function. Generalizing Stiglitz (1982) to a continuum of types, Sachs et al. (2020) consider reforms of arbitrarily nonlinear tax schedules and the optimal taxation in general equilibrium. They demonstrate that increasing tax rates in an initially progressive tax system increases government revenue more with endogenous than with exogenous wages. Therefore, depending on the initial tax code, it may be beneficial to raise tax progressivity. However, the settings of Ales et al. (2015) and Sachs et al. (2020) ignore workers' migration responses. We generalize their results by including the extensive margin.

Rothschild and Scheuer (2013) examine the optimal nonlinear income tax schedule in a multi-sector Roy model with endogenous wages. Trickle-down effects are central for their finding that the optimal tax system is more progressive than in an environment without occupational choice. At first glance, one might think that their two-sector setting nests our two-country economy. The key difference to their paper is that, in our model, two governments set tax policies for each country separately. In Rothschild and Scheuer (2013), only one government chooses the tax schedule for both sectors. The latter, however, is equivalent to the coordinated tax policy setup in our model, which we consider as an extension.

Our continuous-type framework connects the closed economy setup with endogenous wages by Sachs et al. (2020) to the two-country environment of Lehmann et al. (2014) with internationally mobile workers and heterogeneous migration costs. Contrary to the fixed-wage economies in the tax competition literature (e.g., Simula and Trannoy (2010), Bierbrauer, Brett, and Weymark (2013), and Lehmann et al. (2014)), in our environment, workers are imperfect substitutes in producing a composite output good under a CES technology. There are two notable exceptions departing from exogenous wages. One is Tsugawa (2021), who also studies tax competition under endogenous wages but restricts attention to a two-type setup. Another one is Guerreiro, Rebelo, and Teles (2020), demonstrating, in a similar two-type setup, that part of an optimal immigration policy is to use taxes to discourage low-skilled immigration, while encouraging high-skilled immigration.

Altogether, our continuous-type framework nests those of Mirrlees (1971), Lehmann et al. (2014), and Sachs et al. (2020). We derive and quantify the traditional effects occurring in the former papers. Moreover, we discover novel effects that enlighten the interaction between migration and general equilibrium responses and compare their quantitative importance to the traditional effects.

Outline. In Section 2, we solve a discrete K-type Stiglitz (1982) model with general equilibrium and migration responses. Then, we move to a continuous-type Mirrlees (1971) setup and solve for the nonlinear tax incidence and optimal taxation with migration and general equilibrium effects (Section 3). In Section 4, we calibrate our model to the U.S. economy, and compute the optimal tax schedule across different technologies depending on whether migration and general equilibrium effects are present. Section 5 studies the effects of tax coordination under symmetric country sizes and discusses alternative welfare criteria. Section 6 concludes. We relegate all proofs to the Appendix.

2 *K*-Type Model

2.1 Setup

Economic environment. We begin with extending the canonical model of Stiglitz (1982) to a setting where two countries or regions $i = A, B$ compete for internationally mobile workers. As Ales et al. (2015), we go beyond the two-type setting studied in Stiglitz (1982) by considering an arbitrary set of skill or productivity types $\theta \in \Theta = \{1, ..., K\}$. An individual's skill is private information and not observable by the government.

We consider a general class of utility functions, $u(c, l)$, with consumption c, labor supply l, and labor income $y \equiv wl$. Let $u(c, l)$ satisfy the Spence-Mirrlees single-crossing property, $\frac{d}{dw}\frac{-u_l(c,y/w)/w}{u_c(c,y/w)} < 0$, and suppose that labor and consumption are separable (e.g., $u(c, l) = h(c) - v(l)$). We order types such that their equilibrium wages are increasing in types $w_{i,\theta} > w_{i,\theta-1}$ (for a discussion, see Ales et al. (2015)).[4]

Let $n_{i,\theta}$ be the number of natives (born) in country i with skill θ. Denote $N_{i,\theta}$ as country i's equilibrium mass of θ-type workers and $l_{i,\theta}$ as an individual's labor supply. Country i's government taxes labor income according to a nonlinear tax scheme $T_i(y_{i,\theta})$. Consumption of a worker is then given by the after-tax income $c_{i,\theta} = y_{i,\theta} - T_i(y_{i,\theta})$. As we explain later, both the labor supply and the equilibrium population will be endogenous to the tax system.

In each country i, competitive firms produce a single composite output under a constant elasticity of substitution (CES)

$$F_i\left(\{l_{i,\theta}N_{i,\theta}\}_{\theta\in\Theta}\right) = \left[\sum_{\theta\in\Theta} a_{i,\theta}\left(l_{i,\theta}N_{i,\theta}\right)^{\frac{\sigma-1}{\sigma}}\right]^{\frac{\sigma}{\sigma-1}}$$

for some $\sigma \in [0, \infty)$ and $a_{i,\theta} \in \mathbb{R}_+$.[5] Consequently, a worker θ's marginal product pins down her wage rate in that country

$$w_{i,\theta} = a_{i,\theta}\left[l_{i,\theta}N_{i,\theta}/F_i\left(\{l_{i,\theta}N_{i,\theta}\}_{\theta\in\Theta}\right)\right]^{-\frac{1}{\sigma}} \text{ for } \theta \in \Theta, \tag{1}$$

[4] In our continuous-type model, we show that under relatively mild assumptions on the elasticity of labor supply (with respect to the net of tax rate as well as the wage rate) and the marginal tax rate, the ordering of types is without loss of generality ensuring a one-to-one mapping between skills, wages, and incomes (see Appendix B.1, Assumption 1). Therefore, one can interpret a worker's type as her position (e.g., percentile) in the equilibrium wage distribution (e.g., Dustmann et al. (2013)).

[5] In a two-type setup, one can extend the results to any constant-return-to-scale production function.

which she takes as given. Let labor and goods markets clear in each country. Define $\gamma_{i,\theta,\theta} \equiv \frac{\partial \ln\left(w_{i,\theta}\right)}{\partial \ln\left(N_{i,\theta} l_{i,\theta}\right)} < 0$ and $\gamma_{i,\theta,\theta'} \equiv \frac{\partial \ln\left(w_{i,\theta}\right)}{\partial \ln\left(N_{i,\theta'} l_{i,\theta'}\right)} > 0$ as the own- and cross-wage elasticity.

Labor supply. Conditional on living in country i, a worker optimally chooses labor supply $l_{i,\theta}$ to maximize utility $u\left(c_{i,\theta}, l_{i,\theta}\right)$. The worker's first-order condition

$$u_c\left(c_{i,\theta}, l_{i,\theta}\right) w_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right) = -u_l\left(c_{i,\theta}, l_{i,\theta}\right) \tag{2}$$

pins down optimal labor supply.

Migration. As in Lehmann et al. (2014), a worker θ born in country i draws a migration cost m from a conditional density function $G_i\left(m|\theta\right) = \int_0^m g_i\left(x|\theta\right) dx$, accounting for the fact that migration costs may differ between workers (even conditional on skill type). Then, a native in country i, for instance, migrates to country j if and only if $u\left(c_{j,\theta}, l_{j,\theta}\right) - m > u\left(c_{i,\theta}, l_{i,\theta}\right)$. Defining $\Delta_{i,\theta} \equiv u\left(c_{i,\theta}, l_{i,\theta}\right) - u\left(c_{j,\theta}, l_{j,\theta}\right)$, one can derive a country's equilibrium mass of θ-workers as

$$N_{i,\theta} \equiv \rho_i\left(\Delta_{i,\theta}|\theta\right) \equiv \begin{cases} n_{i,\theta} + G_j\left(\Delta_{i,\theta}|\theta\right) n_{j,\theta} & \text{for } \Delta_{i,\theta} \geq 0 \\ \left(1 - G_i\left(-\Delta_{i,\theta}|\theta\right)\right) n_{i,\theta} & \text{for } \Delta_{i,\theta} \leq 0 \end{cases}. \tag{3}$$

Accordingly, denote the semi-elasticity of migration as $\eta_{i,\theta} \equiv \frac{\partial \rho_i\left(\Delta_{i,\theta}|\theta\right)}{\partial \Delta_{i,\theta}} \frac{1}{N_{i,\theta}} \geq 0$.

Government problem. We consider a Nash game between the governments of the two countries. Each government chooses its nonlinear income tax schedule, taking the other country's tax schedule as given and correctly anticipating the migration and labor supply effects from its tax policy. There is no government consumption, and the tax system embraces a purely redistributive motive. As in Simula and Trannoy (2010) and Lehmann et al. (2014), we focus on the most redistributive tax policy and consider a Rawlsian objective function: The government maximizes the average utility of the lowest type. The approach has the advantage that—given the government's objective of redistributing from high and medium types to the lowest type—constraints from mobility become most visible. In addition, one avoids the issue that the aggregate welfare level depends on migration decisions.[6] Formally, country i's government wants to

[6]Defining Pareto weights $\{\psi_{i,\theta}\}_{\theta \in \Theta}$ with $\psi_{i,\theta-1} \geq \psi_{i,\theta}$, this exposition is very similar for a utilitarian objective. A Rawlsian government is a special case where $\psi_{i,1} = 1/N_{i,1}$ and $\psi_{i,\theta} = 0$ for all $\theta \geq 2$. Note that this assumes that $\theta = 1$ is present in both countries at the equilibrium allocations. In Section 5, we consider alternative welfare criteria that have been studied in the literature.

redistribute to the lowest type, and, thus, solves

$$\max_{\{c_{i,\theta},l_{i,\theta},N_{i,\theta}\}_{\theta\in\Theta}} \quad u\left(c_{i,1},l_{i,1}\right) \tag{4}$$

$$\text{subject to} \quad u\left(c_{i,\theta},l_{i,\theta}\right) \geq u\left(c_{i,\theta-1},\frac{l_{i,\theta-1}w_{i,\theta-1}}{w_{i,\theta}}\right) \text{ for } \theta\in\{2,...,K\}, \tag{5}$$

$$\sum_{\theta\in\Theta} N_{i,\theta}c_{i,\theta} \leq F_i\left(\{N_{i,\theta}l_{i,\theta}\}_{\theta\in\Theta}\right), \tag{6}$$

as well as subject to the endogeneity of wages (Equation (1)) and the equilibrium population (Equation (3)), and taking the other country j's allocation as given. Equations (5) and (6) are the workers' incentive constraints and the government budget (no public good provision).

A few comments are in order: Observe that one can omit non-local incentive constraints, excluding cases where a worker θ mimics type $\theta-2, \theta-3, ...$ and $\theta+2, \theta+3, ...$ (see Milgrom and Shannon (1994)). Following Stiglitz (1982) and Ales et al. (2015), we focus on solutions where only the local downward incentive constraints bind. With migration responses, this assumption is not entirely innocuous. We discuss this aspect at the end of this section. Similarly, we focus on solutions where the tax liability increases in types, $T_i\left(y_{i,\theta}\right) \geq T_i\left(y_{i,\theta-1}\right)$, which is natural given the objective of redistributing from the rich to the poor (e.g., set $K=2$). Furthermore, we implicitly assume that governments do not discriminate between natives and immigrant workers in their taxation. Finally, the planner problem is identical to the one in Ales et al. (2015) except for the endogeneity of the equilibrium population (Equation (3)) and the specification of welfare weights to a Rawlsian objective function.

Limitations. We note several limitations to our setup. Firstly, the static nature of our model makes it challenging to confront it with empirically observed migration flows. Secondly, the Nash equilibrium is a rather sophisticated equilibrium concept. Households must form rational expectations about their wage in their country of residence and in the other country. This involves correctly anticipating the migration decisions of everyone else. Moreover, each government takes the other countries' tax scheme as entirely given and, thereby, disregards that any unilateral deviation from the Nash equilibrium tax policy makes the other country's government budget unbalanced, necessarily provoking a readjustment of the tax policy there.

Exogenous technology planner. For comparison, we consider an alternative tax system chosen by an "exogenous technology planner" who ignores migration responses ($\eta_{i,\theta}=0,\ \forall\theta\in\Theta$).

That is, we compute the optimal tax rate of an exogenous technology planner, $T_i^{ex\prime}(y_{i,\theta})$, who ignores migration and maximizes Equation (4) subject to the endogeneity of wages (Equation (1)), the local incentive constraints (5), and the aggregate budget (6). This planner takes as given the number of the population groups $\{N_{i,\theta}\}$ that materializes in the open economy Nash equilibrium, which facilitates comparison with the latter regarding taxation.[7] This notion includes the self-confirming policy equilibrium proposed by Rothschild and Scheuer (2013, 2016), where the exogenous technology planner sets the tax scheme such that it generates outcomes for which it is optimal.[8]

2.2 Optimal Marginal Tax Rates

Optimal top and bottom marginal tax rate. In Proposition 1, we characterize high- and low-skilled workers' optimal (Nash equilibrium) marginal tax rate with migration and compare it to the optimal tax rate of the exogenous technology planner (without migration).

Proposition 1. $a)$ *In the Nash equilibrium of the K-type economy, the optimal marginal tax at the top is higher than in the closed economy, $T_i'(y_{i,K}) > T_i^{ex\prime}(y_{i,K})$.* $b)$ *If the migration semi-elasticity is increasing or constant, $\eta_{i,\theta+1} \geq \eta_{i,\theta}$, and the density of native workers is uniform across types $n_{i,\theta+1} = n_{i,\theta}$, then, in the symmetric Nash equilibrium, the optimal marginal tax at the bottom is lower than in the closed economy, $T_i'(y_{i,1}) < T_i^{ex\prime}(y_{i,1})$.*

Proof. See Appendix A.1. □

Perhaps surprisingly, the optimal marginal tax rate at the top is *higher* with migration than without. This result emerges despite the "migration threat" described in the tax competition literature, where migration leads to *lower* marginal tax rates (e.g., Lehmann et al. (2014)).[9] The reason is that general equilibrium externalities are absent in the existing partial equilibrium models with fixed wages. In general equilibrium, trickle-down forces justify lower top tax

[7] More directly, one may focus on the symmetric Nash equilibrium. Then, we do not have to take a stand on why a government, which does not consider migration responses in its optimization (e.g., in a Mirrlees (1971) benchmark), does not respond to migration flows when observing them.

[8] In Appendix A.2, we derive closed-form expressions for $T_i^{ex\prime}(y_{i,\theta})$ in a simplified setting with two types, a Cobb-Douglas technology, a linear consumption utility, an isoelastic disutility of labor, and symmetric countries. Then, there is no endogeneity of the right-hand side variables in the optimal tax formula.

[9] The possibility that migration opportunities of the rich may increase tax rates has been noted in other situations (absent of general equilibrium wage effects), such as endogenous land quality (see Glazer, Kanniainen, and Poutvaara (2008)).

rates relative to an economy with fixed wages because a tax cut at the top raises low-skilled wages and lowers high-skilled wages (predistribution). With labor migration, these general equilibrium forces may still call for a lower marginal tax of high-skilled workers (compared to the partial equilibrium) but less relative to an economy without migration. The intuition is that a reduction in high-skilled wages causes an outflow of these workers marginally raising their wages such that the predistribution rationale is self-limiting. In that sense, trickle-down forces are partly offset by labor migration. At the very top, this force overturns the migration threat, as we make more transparent in the continuous-type model in the next section.

For the result on the bottom tax rate, we need additional assumptions. Firstly, we rule out that the migration semi-elasticity decreases for some parts of the income distribution.[10] Secondly, we assume that the number of each type in the native population is constant (uniform density), which is not very restrictive.[11] Thirdly, we focus on the symmetric Nash equilibrium.

With these assumptions the bottom tax rate is lower in the open than in the closed economy. The intuition is the same as the one that calls for a lower marginal subsidy at the top. In response to a lower marginal tax rate, low-skilled workers' labor supply rises. On the one hand, this leads to a decline in low-skilled workers' wage rates. But, on the other hand, due to the complementarity of labor, the wages of high-skilled mobile workers increase. Through this wage channel, the lower marginal tax at the bottom amplifies pre-tax wage inequality in the respective country trying to attract high-skilled workers. Altogether, the K-type model suggests that the presence general equilibrium effects partly offsets the "migration threat" described in Lehmann et al. (2014).[12]

A note of caution. The presence of international migration opportunities makes optimal tax codes more redistributive at the ends of the type distribution in terms of marginal tax rates. Nonetheless, the effect of migration on tax payments (or transfers, respectively) and thus average tax burdens could be the opposite. To understand this ambiguity, consider the case with only two types, $K = 2$. Starting from the optimal consumption allocation without migration and holding labor supply fixed, a revenue-neutral reduction in lump-sum payments to low-skilled

[10] In the next section, we also consider a decreasing semi-elasticity.

[11] For instance, one can interpret a worker's type θ as her percentile rank in the income distribution, in which case $n_{i,\theta} = \frac{1}{K}$.

[12] We also derive results about the Nash equilibrium tax rates of middle incomes. In our working paper version, we demonstrate that the comparison with the taxes set by the exogenous technology planner depends on how much the specific middle-income type contributes to wage inequality (see Janeba and Schulz (2021)).

workers leads to high-skill immigration and low-skill emigration. The government can use the resulting fiscal surplus for transfers to low-skilled workers leading to a welfare improvement. However, this line of reasoning is not complete to prove lower tax burdens on high-skilled workers because wages and labor supplies also change, thereby affecting tax payments. Therefore, one cannot infer from the results on the responses of marginal tax rates to the reaction of average tax rates and, thus, the level of redistribution.

Moreover, we have focused on solutions where only the local downward incentive constraints bind. This assumption makes our results directly comparable to those in the literature on optimal taxation in general equilibrium (e.g., Stiglitz (1982) and Ales et al. (2015)). In the presence of migration, however, the assumption is debatable because sufficiently increasing migration semi-elasticities may imply decreasing tax payments at the very top (see Lehmann et al. (2014)). In such a situation, local upward incentive constraints would bind. This issue is local as it would only concern workers at the very top—in our model the worker $\theta = K$. Notwithstanding, our solution does not consider the possibility of binding local upward incentive constraints. Both observations motivate the continuous-type framework in the next section.

3 Continuous-Type Model

In the K-type model considered above, we could only obtain sharp results about the optimal top and bottom tax rates. In the following, we develop a standard (Mirrleesian) continuous-type economy with migration and general equilibrium wage responses. We follow the standard procedure of first analyzing the nonlinear tax incidence and, then, solving for an optimal arbitrarily nonlinear tax scheme. Later, we use the framework to simulate for the optimal tax scheme. Our approach allows us to directly relate to the literature on optimal taxation in general equilibrium (Sachs et al. (2020)), and to provide a connection to the one on optimal taxation with migration in partial equilibrium (Lehmann et al. (2014)).

3.1 Setup

We retain the two-country framework. In contrast to Section 2, we assume that, in country $i \in \{A, B\}$, there is a continuum of workers $\theta \in \Theta = [0, 1] \sim F_i(\theta)$. Without loss of generality, a worker's type θ can be interpreted as her position in the wage distribution (see

Assumption 1 in Appendix B.1). As before, each worker's utility function satisfies the Spence-Mirrlees single-crossing property. However, we now abstract from income effects on labor supply. For instance, the utility function may represent GHH-preferences $u(c,l) = u(c - v(l))$ (see Greenwood, Hercowitz, and Huffman (1988)), where $v(l)$ denotes an increasing, convex disutility from labor. For a given wage rate $w_{i,\theta}$, a worker's first-order condition

$$w_{i,\theta}\left(1 - T_i'(y_{i,\theta})\right) = v'(l_{i,\theta}) \tag{7}$$

pins down labor supply. Following, e.g., Scheuer and Werning (2017) and Jacquet and Lehmann (2021), we define the reduced-form elasticities of optimal labor supply along the nonlinear budget line with respect to the marginal income tax rate and the wage rate, respectively, as

$$\varepsilon_{i,\theta}^{l,(1-T')} \equiv \frac{\partial \ln l_{i,\theta}}{\partial \ln\left(1 - T_i'(y_{i,\theta})\right)} = \frac{\varepsilon_{i,\theta}^{l,(1-\tau)}}{1 + p_i(y_{i,\theta})\, \varepsilon_{i,\theta}^{l,(1-\tau)}}$$

and

$$\varepsilon_{i,\theta}^{l,w} \equiv \left(1 - p_i(y_{i,\theta})\right) \varepsilon_{i,\theta}^{l,(1-T')},$$

where $\varepsilon_{i,\theta}^{l,(1-\tau)} \equiv \frac{\partial \ln l_{i,\theta}}{\partial \ln\left(1-\tau_{i,\theta}\right)}$ denotes the elasticity along the linear budget line, and $p_i(y_{i,\theta}) \equiv -\frac{\partial \ln\left(1-T_i'(y_{i,\theta})\right)}{\partial \ln y_{i,\theta}}$ defines the local rate of progressivity of the income tax scheme.

As before and in Lehmann et al. (2014), a conditional migration cost distribution $G_i(m|\theta)$ gives rise to an endogenous equilibrium population distribution $N_{i,\theta} \equiv \rho_i(\Delta_{i,\theta}|\theta)$ for each worker θ and country i, where $\Delta_{i,\theta}$ denotes the utility difference from living in country i relative to country j (before migration cost). Again, denote $\eta_{i,\theta} \equiv \frac{\partial \rho_i\left(\Delta_{i,\theta}|\theta\right)}{\partial \Delta_{i,\theta}} \frac{1}{N_{i,\theta}}$ as the semi-elasticity of migration. In each country i, a mass-one continuum of identical firms produces a single final output good using the labor of each worker type and a CES technology

$$F_i\left(\{l_{i,\theta} N_{i,\theta}\}_{\theta\in\Theta}\right) = \left[\int_{\theta\in\Theta} a_{i,\theta} \left(l_{i,\theta} N_{i,\theta}\right)^{\frac{\sigma-1}{\sigma}} d\theta\right]^{\frac{\sigma}{\sigma-1}}$$

for some $\sigma \geq 0$ and $a_{i,\theta} \in \mathbb{R}_+$. Firms earn zero profits, and, as in the discrete-type setting, workers' wages are equal to the respective marginal product of labor

$$w_{i,\theta} = a_{i,\theta}\left[l_{i,\theta} N_{i,\theta} / F_i\left(\{l_{i,\theta} N_{i,\theta}\}_{\theta\in\Theta}\right)\right]^{-\frac{1}{\sigma}} \text{ for } \theta \in \Theta. \tag{8}$$

For this CES technology, $\sigma = 0$, $\sigma = 1$, and $\sigma = \infty$ correspond to Leontieff, Cobb-Douglas, and exogenous-wage technologies.

As in the K-type setting above, we focus on the optimal Rawlsian tax system (no public good provision). In each country i, the government chooses the tax schedule to maximize the average indirect utility of the lowest worker type subject to the aggregate budget constraint

$$\max_{\{T_i(y_{i,\theta})\}_{\theta\in\Theta}} \mathcal{U}_{i,0} \quad \text{subject to} \quad \mathcal{R}_i \equiv \int_{\theta\in\Theta} T_i\left(y_{i,\theta}\right) N_{i,\theta} d\theta \geq 0, \tag{9}$$

as well as subject to the equilibrium supply and demand of labor (Equations (7) and (8)), the endogenous migration responses (Equation (3)), and taking as given the other country's tax system.

3.2 Nonlinear Tax Incidence

Before deriving the optimal tax system, we characterize the incidence of reforming the tax system. Starting from an arbitrary initial tax code, the government can implement an arbitrarily nonlinear tax reform $\hat{T}_i\left(y_{i,\theta}\right)$, whereby the statutory tax payment at income $y_{i,\theta}$ changes from $T_i\left(y_{i,\theta}\right)$ to $T_i\left(y_{i,\theta}\right) + \mu\hat{T}_i\left(y_{i,\theta}\right)$.[13] This approach allows us to gain intuition about the underlying economic forces that drive the optimal choice of the tax system. For a detailed analysis of tax incidence, we refer to Appendix B.2 and B.3. As a byproduct, we obtain the planner's first-order condition that characterizes the solution to the taxation problem (Appendix B.4).

Individual incidence of tax reforms. We start our analysis of tax incidence by describing the individual responses to a marginal tax reform $\hat{T}_i$ (with $\mu \to 0$). We denote $\frac{\hat{l}_{i,\theta}}{l_{i,\theta}}$, $\frac{\hat{w}_{i,\theta}}{w_{i,\theta}}$, and $\frac{\hat{N}_{i,\theta}}{N_{i,\theta}}$ as the percentage change of an individual's labor supply, wage rate, and equilibrium population. Absent of income effects, a worker's labor supply responds in general equilibrium in two respects: directly through the behavioral effect and indirectly due to the adjustment in her wage

$$\frac{\hat{l}_{i,\theta}}{l_{i,\theta}} = \varepsilon_{i,\theta}^{l,(1-T')} \frac{-\hat{T}_i'\left(y_{i,\theta}\right)}{1-T_i'\left(y_{i,\theta}\right)} + \varepsilon_{i,\theta}^{l,w} \frac{\hat{w}_{i,\theta}}{w_{i,\theta}}. \tag{10}$$

Similarly, we perturb the equilibrium population to show that the response of the equilibrium

[13] Thus, the map $\hat{T}_i$ defines the (infinite-dimensional) direction of the tax reform, while the scalar μ parametrizes its size. We assume that the tax reforms $\hat{T}_i$ that the government can implement belong to the Banach space of functions that are continuously differentiable, with a bounded first derivative.

population consists of a direct (mechanical) effect and a wage effect

$$\frac{\hat{N}_{i,\theta}}{N_{i,\theta}} = -\hat{T}_i\left(y_{i,\theta}\right)\eta_{i,\theta} + \eta_{i,\theta}y_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right)\frac{\hat{w}_{i,\theta}}{w_{i,\theta}}. \tag{11}$$

By the envelope theorem, behavioral effects play no first-order role for the equilibrium population.

To determine the impact on the labor supply and the equilibrium population, we derive the incidence on wages by perturbing the wage equation (8)

$$\frac{\hat{w}_{i,\theta}}{w_{i,\theta}} = -\frac{1}{\sigma}\left(\frac{\hat{l}_{i,\theta}}{l_{i,\theta}} + \frac{\hat{N}_{i,\theta}}{N_{i,\theta}}\right) + \frac{1}{\sigma}\frac{\left[\int_{\theta\in\Theta} a_{i,\theta}\left(\frac{\hat{l}_{i,\theta}}{l_{i,\theta}} + \frac{\hat{N}_{i,\theta}}{N_{i,\theta}}\right)\left(l_{i,\theta}N_{i,\theta}\right)^{\frac{\sigma-1}{\sigma}}d\theta\right]}{\left[\int_{\theta\in\Theta} a_{i,\theta}\left(l_{i,\theta}N_{i,\theta}\right)^{\frac{\sigma-1}{\sigma}}d\theta\right]}. \tag{12}$$

By solving this system of three equations (10), (11), and (12), the wage incidence can be written in closed form[14]

$$\begin{aligned}\frac{\hat{w}_{i,\theta}}{w_{i,\theta}} = &-\frac{1}{\sigma}\frac{1}{1+\frac{1}{\sigma}\left(\varepsilon_{i,\theta}^{l,w} + \eta_{i,\theta}y_{i,\theta}\left(1-T'\left(y_{i,\theta}\right)\right)\right)}\left\{\frac{\hat{l}_{i,\theta}^{PE}}{l_{i,\theta}} + \frac{\hat{N}_{i,\theta}^{PE}}{N_{i,\theta}}\right.\\ &\left.-\frac{\int_{\theta\in\Theta}\left(\frac{dl_{i,\theta}^{PE}}{l_{i,\theta}} + \frac{dN_{i,\theta}^{PE}}{N_{i,\theta}}\right)y_{i,\theta}N_{i,\theta}/\left[1+\frac{1}{\sigma}\left(\varepsilon_{i,\theta}^{l,w} + \eta_{i,\theta}y_{i,\theta}\left(1-T_i'\left(y_{i,\theta}\right)\right)\right)\right]d\theta}{\int_{\theta\in\Theta} y_{i,\theta}N_{i,\theta}/\left[1+\frac{1}{\sigma}\left(\varepsilon_{i,\theta}^{l,w} + \eta_{i,\theta}y_{i,\theta}\left(1-T_i'\left(y_{i,\theta}\right)\right)\right)\right]d\theta}\right\},\end{aligned}$$

where we denote $\frac{\hat{l}_{i,\theta}^{PE}}{l_{i,\theta}} + \frac{\hat{N}_{i,\theta}^{PE}}{N_{i,\theta}} \equiv \varepsilon_{i,\theta}^{l,(1-T')}\frac{-\hat{T}_i'\left(y_{i,\theta}\right)}{1-T_i'\left(y_{i,\theta}\right)} - \hat{T}_i\left(y_{i,\theta}\right)\eta_{i,\theta}$ as the partial equilibrium labor supply and migration adjustment. Thus, the wage responses depend on the presence of migration responses. To make this more transparent, we set $\eta_{i,\theta} = 0$ for any θ in the expression for the wage incidence and define $\frac{\hat{w}_{i,\theta}^{STW}}{w_{i,\theta}}$ as the wage change absent of migration responses (see Sachs et al. (2020), abbreviated as "STW" in the following).

Aggregate incidence of tax reforms. We start our analysis of aggregate tax incidence with a decomposition.

Lemma 1. *The incidence of a tax reform $\hat{T}_i$ of any initial tax schedule T_i on government revenues can be decomposed into a mechanical effect, a behavioral effect, a mechanical migration*

[14]Plugging Equation (12) into the sum of (10) and (11) gives an inhomogeneous Fredholm integral equation of the second kind that one can solve using standard techniques.

effect, a wage effect, a migration effect on wages, and a wage effect on migration:

$$\hat{\mathcal{R}}_i = \mathcal{ME}_i + \mathcal{BE}_i + \mathcal{MME}_i + \mathcal{WE}_i + \mathcal{MEW}_i + \mathcal{WEM}_i, \tag{13}$$

where the components are defined by the following Equations (14)–(19).

The first term captures the direct effect of a small rise in T_i (i.e., higher tax payments) on aggregate tax revenues

$$\mathcal{ME}_i \equiv \int_{\theta\in\Theta} \hat{T}_i\left(y_{i,\theta}\right) N_{i,\theta} d\theta. \tag{14}$$

In the literature (e.g., Saez (2001)), this effect is referred to as the *mechanical effect*. The second term collects labor supply effects. A higher marginal tax rate reduces workers' incentives to supply labor which is commonly labeled as the *behavioral effect*

$$\mathcal{BE}_i \equiv \int_{\theta\in\Theta} \frac{\hat{l}_{i,\theta}^{PE}}{l_{i,\theta}} y_{i,\theta} T_i'\left(y_{i,\theta}\right) N_{i,\theta} d\theta. \tag{15}$$

The combination of the mechanical and the behavioral effect leads to a Diamond-Saez formula for the optimal tax system in partial equilibrium without migration.

The third term captures the mechanical effect of changing the tax code on the equilibrium population

$$\mathcal{MME}_i \equiv -\int_{\theta\in\Theta} T_i\left(y_{i,\theta}\right) \hat{T}_i\left(y_{i,\theta}\right) \eta_{i,\theta} N_{i,\theta} d\theta. \tag{16}$$

For instance, a rise in tax payments leads to labor emigration and, thus, lower tax revenues. This negative impact of labor mobility limiting a government's ability to levy high taxes is typically referred to as the threat of migration (e.g., Lehmann et al. (2014)).

In general equilibrium, not only labor supply but also wages (that is, labor demand) respond to tax reforms affecting aggregate tax revenues

$$\mathcal{GE}_i \equiv \int_{\theta\in\Theta} \left(1+\varepsilon_{i,\theta}^{l,w}\right) \frac{\hat{w}_{i,\theta}}{w_{i,\theta}} y_{i,\theta} T_i'\left(y_{i,\theta}\right) N_{i,\theta} d\theta.$$

The general equilibrium effect, $\mathcal{GE}_i = \mathcal{WE}_i + \mathcal{MEW}_i$, can be decomposed into a standard wage effect that captures the revenue effect of wage changes absent of migration (see Sachs et al. (2020))

$$\mathcal{WE}_i \equiv \int_{\theta\in\Theta} \left(1+\varepsilon_{i,\theta}^{l,w}\right) \frac{\hat{w}_{i,\theta}^{STW}}{w_{i,\theta}} y_{i,\theta} T_i'\left(y_{i,\theta}\right) N_{i,\theta} d\theta \tag{17}$$

and a novel *migration effect on wages*

$$\mathcal{MEW}_i \equiv \int_{\theta\in\Theta} \left(1+\varepsilon_{i,\theta}^{l,w}\right)\left(\frac{\hat{w}_{i,\theta}}{w_{i,\theta}} - \frac{\hat{w}_{i,\theta}^{STW}}{w_{i,\theta}}\right) y_{i,\theta} T_i'\left(y_{i,\theta}\right) N_{i,\theta} d\theta. \tag{18}$$

As shown by Stiglitz (1982) and Sachs et al. (2020), the standard wage response $\mathcal{WE}_i$ calls for a less progressive tax code when the government optimally chooses an arbitrarily nonlinear tax schedule. The intuition is that a government can lower the inequality in pre-tax wages by taxing the poor more and the rich less. This tax-induced reduction in wage inequality is called predistribution.

However, here the presence of migration responses adjusts the wage responses. The aggregate effect is summarized by the migration effect on wages $\mathcal{MEW}_i$. For instance, a rise in top tax rates triggers an emigration response by high-skilled workers. As a result, their aggregate labor supply marginally declines and pre-tax wage inequality goes up. This inequality rise works against predistribution but expands the tax base.[15]

Finally, the last term captures a novel *wage effect on migration* that collects the first-order effects of wage changes on the equilibrium population

$$\mathcal{WEM}_i \equiv \int_{\theta\in\Theta} T_i\left(y_{i,\theta}\right) \frac{\hat{w}_{i,\theta}}{w_{i,\theta}} y_{i,\theta}\left(1-T_i'\left(y_{i,\theta}\right)\right)\eta_{i,\theta} N_{i,\theta} d\theta. \tag{19}$$

As the migration effect on wages partly offsets predistribution, the wage effect on migration works against the standard migration threat. For instance, by raising taxes at the top and, hence, amplifying pre-tax wage inequality (lower high-skilled labor supply), a government can attract high-skilled workers and raise more tax revenues.

We conclude the incidence analysis by deriving explicit expressions for the migration effect on wages and the wage effect on migration using a specific tax reform and model parametrization. In particular, we consider the class of elementary tax reforms that raises tax payments of anyone above income level y^* and yields a mechanical effect on government revenues of 1\$ (see Saez (2001)).[16] We assume an isoelastic disutility of labor (constant labor supply elasticities), $v\left(l\right) = \frac{l^{1+1/e}}{1+1/e}$, an initial tax schedule that has a constant rate of progressivity (CRP),

[15]The idea that, in the presence of labor complementarities, some workers' immigration affects others' wages and, thereby, raises tax revenues is similar to Colas and Sachs (2020) who quantify the indirect (fiscal) benefit of low-skilled immigration on high-skilled workers' wages.

[16]Formally, $\hat{T}_i\left(y_{i,\theta}\right) = \frac{\mathbb{1}\left(y_{i,\theta}\geq y^*\right)}{1-\int_{\theta\leq\theta^*} N_{i,\theta} d\theta}$ and $\hat{T}_i'\left(y_{i,\theta}\right) = \frac{\delta\left(y_{i,\theta}-y^*\right)}{1-\int_{\theta\leq\theta^*} N_{i,\theta} d\theta}$.

$T_i(y) = \frac{1-\tau_i}{1-p_i} y^{1-p_i}$,[17] and a constant migration elasticity $\nu_i = \eta_{i,\theta} c_{i,\theta}$. Under these assumptions, Corollary 1 further characterizes the two novel effects identified in Lemma 1.

Corollary 1. *Suppose the disutility of labor is isoleastic, the initial tax schedule is CRP, and the migration elasticities are constant. Consider an elementary tax reform at y^*. Then,*

$$\mathcal{WEM}_i(y^*) = t_{1,i} \mathcal{Z}_{1,i}(y^*) + t_{2,i} \mathcal{Z}_{2,i}(y^*)$$

and

$$\mathcal{MEW}_i(y^*) = -t_{3,i} \mathcal{Z}_{1,i}(y^*) + t_{4,i} \mathcal{Z}_{2,i}(y^*),$$

where

$$\mathcal{Z}_{1,i}(y^*) \equiv \frac{y^* N_{y^*}}{1 - \int_{y' \leq y^*} N_{i,y'} dy'} \int_y \frac{T'(y^*) - T'(y)}{1 - T'(y^*)} \frac{y N_y dy}{\int_{y'} y' N_{y'} dy'},$$

$$\mathcal{Z}_{2,i}(y^*) \equiv \int_{y > y^*} \int_{y'} \frac{T'(y) - T'(y')}{1 - T'(y)} \frac{y' N_{y'} dy'}{\int_{y''} y'' N_{y''} dy''} \frac{N_y dy}{1 - \int_{y'' \leq y^*} N_{i,y''} dy''},$$

and $t_{1,i}$, $t_{2,i}$, $t_{3,i}$, and $t_{4,i}$ are positive constants defined in Appendix B.3. Omitting higher-order effects ($\nu_i/\sigma^2 \to 0$ and $\nu_i^2/\sigma \to 0$), $\mathcal{WEM}_i(y^*) \to t_{1,i} \mathcal{Z}_{1,i}(y^*)$ and $\mathcal{MEW}_i(y^*) \to t_{4,i} \mathcal{Z}_{2,i}(y^*)$.

An elementary tax reform changes the tax code in two respects. Firstly, it raises the marginal tax rate of the worker who just earns y^* in the initial equilibrium. The worker responds by reducing the labor supply, changing the entire wage distribution and triggering migration, proportional to $\mathcal{Z}_{1,i}(y^*)$. Thus, this term connects to the wage effect on migration. Also, observe the close similarity of $\mathcal{Z}_{1,i}(y^*)$ to the expressions in Corollary 4 in Sachs et al. (2020). Secondly, the reform raises everyone's tax bill with income above y^*. As a result, these workers emigrate, causing an adjustment of the wage distribution, measured by $\mathcal{Z}_{2,i}(y^*)$. This term is therefore closely related to the migration effect on wages. In summing the respective effects over $y > y^*$, its structure resembles the mechanical migration effect discovered by Lehmann et al. (2014).

One can observe that, under a progressive tax code with increasing marginal tax rates ($p_i > 0$), both terms $\mathcal{WEM}_i(y^*) \to t_{1,i} \mathcal{Z}_{1,i}(y^*)$ and $\mathcal{MEW}_i(y^*) \to t_{4,i} \mathcal{Z}_{2,i}(y^*)$ are positive for a high income y^*. Accordingly, the interaction of migration and general equilibrium responses

[17] By specifying CRP tax functions (developed in Feldstein (1969), Persson (1983), and Benabou (2000)), we follow Heathcote, Storesletten, and Violante (2017), who document that the current U.S. tax code is close to CRP. Heathcote and Tsujiyama (2021a) and Heathcote and Tsujiyama (2021b) study the CRP tax scheme in a Mirrlees (1971) model showing that it approximates well the full optimum.

provides an incentive to raise top taxes. We conclude that both novel effects tend to weaken predistribution and work against the threat of migration that calls for lower top tax rates in response to a rise in labor mobility.

3.3 Optimal Taxation

Having analyzed the traditional and novel effects of migration and general equilibrium forces on tax revenues around an arbitrary initial tax code, we now turn to a characterization of the optimal nonlinear (Rawlsian) tax schedule.[18] By utilizing the property that under an optimal tax system, the aggregate welfare effects of any tax reform are equal to zero, we obtain a Diamond-Saez formula for the optimal tax system (see Diamond (1998) and Saez (2001)).

Proposition 2. *In the Nash equilibrium of the continuous-type economy, the optimal marginal tax is given by*

$$\frac{T_i'(y^*)}{1-T_i'(y^*)} = \mathcal{A}_i(y^*)\,\mathcal{B}_i(y^*)\,\mathcal{C}_i(y^*) - \mathcal{D}_i(y^*), \tag{20}$$

where

$$\mathcal{A}_i(y^*) \equiv \frac{1+\frac{1}{\sigma}\left[\varepsilon_{i,y^*}^{l,w} + y^*\left(1-T_i'(y^*)\right)\eta_{i,y^*}\right]}{\varepsilon_{i,y^*}^{l,(1-T')}},$$

$$\mathcal{B}_i(y^*) \equiv \frac{1-\int_{y'\le y^*} N_{i,y'}\,dy'}{y^* N_{i,y^*}},$$

$$\mathcal{C}_i(y^*) \equiv \int_{y^*}^{\overline{y}} \left[1-\eta_{i,y}T_i(y) - \frac{1}{\sigma}y\eta_{i,y}\frac{\left(1-\eta_{i,y}T_i(y)\right)\left(1-T_i'(y)\right)-\varepsilon_{i,y}^{l,w}T_i'(y)}{1+\frac{1}{\sigma}\left[\varepsilon_{i,y}^{l,w}+y\left(1-T_i'(y)\right)\eta_{i,y}\right]}\right]\frac{N_y\,dy}{1-\int_{y'\le y^*}N_{y'}\,dy'},$$

and

$$\mathcal{D}_i(y^*) \equiv \frac{1-y^*\eta_{i,y^*}\left[T_i(y^*)/y^* - T_i'(y^*)\right]}{\sigma}.$$

Proof. See Appendix B.4. □

Equation (20) extends the formulas in Lehmann et al. (2014) and Sachs et al. (2020) by including the interaction of migration and general equilibrium effects. In particular, for $\eta_{i,y} = 0,\ \forall y > 0$, the formula collapses to the one in Sachs et al. (2020). Similarly, set $\sigma = \infty$ (i.e., exogenous wages) to obtain the optimal tax in Lehmann et al. (2014). The interaction adjusts the

[18] In Section 5, we discuss alternative welfare criteria.

well-known ABC-formulas for optimal taxation in three respects. Firstly, the standard inverse elasticity rule, captured by $\mathcal{A}_i(y^*)$, includes wage effects. When the government raises the tax rate of all workers with income y^*, these workers' wages rise due to labor supply adjustments (see Sachs et al. (2020)) and, now, also migration responses. This own-wage effect mitigates the behavioral response to marginal taxes.

The second adjustment regards the tax-revenue effects of an increase in tax payments above income y^*, summarized in $\mathcal{C}_i(y^*)$. The term $\mathcal{C}_i(y^*)$ includes the mechanical change in tax revenues (for a fixed wage and population distribution) and the usual negative out-migration response to the tax bill rise (migration threat, see Lehmann et al. (2014)). However, when wages change endogenously, one needs to include the described migration effect on wages and the wage effect on migration that account for the migration-induced adjustment in the wage response and the migration response to a wage change, respectively.

The wage effect on migration and the migration effect on wages are not only present above but also at income y^* . Therefore, these two novel effects are also present in the $\mathcal{D}_i(y^*)$-term, which is the third adjustment. The term captures the direct impact of a wage change at y^* on tax revenues. On the one hand, a higher wage at y^* triggers an in-migration response of these workers. The resulting positive revenue effect is measured in terms of the average tax rate. On the other hand, this effect is self-limiting because workers' in-migration itself lowers their wage rate, causing a marginal decline in revenues measured by the marginal tax rate.

Finally, observe how Equation (20) nests Proposition 1. In our K-type economy, the income distribution is bounded by the top type $\theta = K$ (thus, $\overline{y} < \infty$). Therefore, $\mathcal{C}_i(y^*) \to 0$ as $y^* \to \overline{y}$ and the tax rate converges to

$$\frac{T_i'(y^*)}{1 - T_i'(y^*)} \to -\mathcal{D}_i(\overline{y}) = -\frac{1 - \overline{y}\eta_{i,\overline{y}}\left[T_i(\overline{y})/\overline{y} - T_i'(\overline{y})\right]}{\sigma}.$$

In our later simulation, we can observe that $T_i(\overline{y})/\overline{y} > T_i'(\overline{y})$ since the top marginal tax rate either converges to a negative value or it is small compared to the average tax rate. As a result, the presence of migration upwards adjusts the top tax rate (unlike in Lehmann et al. (2014)). Depending on the migration semi-elasticity at the top, it may even be that the marginal tax rate at the top is positive, overturning the usual general equilibrium rationale. In this case, the presence of general equilibrium effects leads to a higher top tax rate (unlike in Sachs et al. (2020)).

4 Numerical Simulations

Our analysis of the discrete and continuous type models provides insights into the qualitative properties of the optimal tax schedule and the channels through which forces operate. Further results are hard to obtain because the formula for the optimal income tax scheme is an integro-differential equation that is not trivial to solve. All right-hand side terms are endogenous objects that depend on the distribution of migration semi-elasticities and the tax system at which they are evaluated.[19] Therefore, we now turn to numerical simulations to illustrate how migration responses and endogenous wages, respectively, affect the optimal tax schedule.

4.1 Calibration

We calibrate our continuous-type model to match empirical facts from the U.S. We assume that the current U.S. tax schedule can be approximated by a CRP tax schedule with parameters $p_i = 0.181$ and $\tau_i = -3.56$ (Sachs et al. (2020), Heathcote et al. (2017)). Moreover, to match the current U.S. labor income distribution, we proceed as in Sachs et al. (2020): Let earnings below $\$150\,000$ be log-normally distributed with mean 10 and variance 0.95. Above $\$150\,000$, we append a Pareto distribution with a tail parameter that decreases from 2.5 to 1.5 (above $\$250\,000$). We consider the symmetric Nash equilibrium, in which case we do not have to calibrate positive migration flows between the U.S. and the rest of the world. We set the Frisch elasticity of labor supply to $e = 0.33$ (Chetty (2012)).

In the calibration, we follow Sachs et al. (2020) who extend the approach of Saez (2001): We use the empirical U.S. income distribution and the individual first-order conditions to infer the wage distribution (Saez (2001)) and, for a given elasticity of substitution, to back out the underlying productivity distribution (Sachs et al. (2020)), $\{a_{i,\theta}\}_{\theta\in\Theta}$. Accordingly, a worker's position in the wage distribution defines her skill level, as in Dustmann et al. (2013), who study the impact of migration on wages. We set the elasticity of substitution to $\sigma = 2.0$ (Katz and Murphy (1992), Card and Lemieux (2001), Card (2009)).[20]

We assume an exponential migration cost distribution $G_i\left(m|\theta\right) = 1 - e^{-\delta_{i,\theta} m}$. Since

[19]The latter issue is particularly important in the context of general equilibrium effects (see Section 3.2 in Sachs et al. (2020) for more details).

[20]Card (2009) reports an upper bound of $\sigma = 2.5$ and a lower bound of $\sigma = 1.5$. Thus, this value is considered to be in the middle of the likely values (Card (2009)). In the literature, there also exist values outside of this range, e.g., $\sigma = 0.6$ (Dustmann et al. (2013)) and $\sigma = 3.1$ (Heathcote et al. (2017)).

the migration semi-elasticity is equal to the hazard rate of the migration cost distribution, $\eta_{i,\theta} = \frac{g_i(\Delta_{i,\theta}|\theta)}{1-G_i(\Delta_{i,\theta}|\theta)} = \delta_{i,\theta}$, in this specification, the semi-elasticity is a model primitive.[21] We are not aware of any direct evidence of the population-wide distribution of the migration semi-elasticity. In principle, it may be possible to calibrate semi-elasticities from the distribution of migration elasticities and a correctly specified tax-and-transfer system. Unfortunately, evidence of workers' mobility to taxes other than at the top of the income distribution is also scarce (see Kleven et al. (2020)).[22] At best, there is some established evidence that high-skilled workers respond more sensitively than low-skilled ones, e.g., to local labor market shifts (see Bound and Holzer (2000)). However, there is no consensus about the causal drivers of this fact. The literature has identified important confounders, such as the level of local amenities (see Diamond (2016)), means-tested government transfers (see Notowidigdo (2020)), immigrant status, and gender differences (see Piyapromdee (2021)). In light of missing robust evidence on this key empirical object, we return to the specification in Lehmann et al. (2014) and consider three different cases about the shape of the migration semi-elasticity (increasing, constant, and decreasing). Since the empirical evidence somewhat suggests that migration forces are primarily present at the top,[23] we focus on an increasing semi-elasticity of migration in the main text and relegate the other cases to the Appendix C. All three cases give rise to an increasing migration elasticity.[24]

We choose the semi-elasticities such that the average migration elasticity of the top 1% of the income distribution is constant $\mathbb{E}_{\theta \geq 0.99}(\nu_{i,\theta}) = \nu_{top}$ (Lehmann et al. (2014)). Following Kleven et al. (2020), we choose a conservative value for the top migration elasticity: $\nu_{top} = 0.25$.[25]

[21]In the symmetric Nash equilibrium, the semi-elasticity is a model primitive for any cost distribution because utility differences are zero and, thus, $\eta_{i,\theta} = \frac{g_i(0|\theta)}{1-G_i(0|\theta)}$. Therefore, following the structural labor literature (e.g., Diamond (2016), Lessem (2018), Piyapromdee (2021)), one could consider location taste shocks that follow a Type 1 Extreme Value distribution, in which case the migration cost (taste shock difference) is distributed logistic. This would lead to a constant migration semi-elasticity in a given worker group.

[22]Relative to high-skilled migration, the empirical estimation of tax-induced low-skilled migration faces two critical challenges. First, the average tax rate, which determines their migration incentives, is an endogenous object and, beyond that, hard to measure due to the presence of social insurance and welfare programs. It cannot be proxied by a small set of exogenous variables, such as the marginal top tax rate in the case of high-skilled migration. Second, tax variation at the bottom is typically associated with other policy changes that may affect migration choices.

[23]Otherwise, one could observe significantly positive (albeit small) migration elasticities in lower parts of the income distribution.

[24]We also considered alternative cases where the migration elasticity is constant for most incomes or all migration responses are concentrated among the middle class. These cases imply substantially declining semi-elasticities. We consider the policy implications of a decreasing semi-elasticity below.

[25]Other papers, such as Kleven, Landais, Saez, and Schultz (2014), who analyze the effects of a tax rebate for

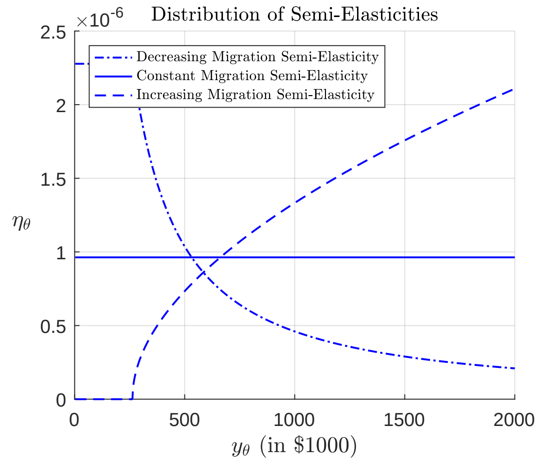


Figure 1: Left Panel: Calibrated Migration Semi-Elasticities; Right Panel: Calibrated Migration Elasticities

We assume that the migration semi-elasticity is zero up to the top 1% and then increasing. As described, in the Appendix, we also consider cases where the semi-elasticity of migration is a constant and where it is decreasing over the population, following Lehmann et al. (2014). The results remain qualitatively unchanged. In Figure 1, we depict the calibrated migration elasticities and semi-elasticities. In our baseline economy with an increasing semi-elasticity, the calibrated population-wide migration elasticity is 0.003 and, under a constant (decreasing) semi-elasticity, it is 0.010 (0.020). In the following, we switch on and off migration and general equilibrium effects ($\eta_{i,\theta} = 0, \ \forall\theta \in [0,1]$, and $\sigma = \infty$, respectively) to compare optimal tax schemes across different environments.

4.2 Presence of Migration

We start by studying the presence of migration forces, in the spirit of Section 2. In the left panel of Figure 2, we depict the optimal marginal tax rates in general equilibrium, turning on and off migration. More formally, we simulate optimal tax schemes when $\eta_{i,\theta} = 0$ for any $\theta \in [0,1]$ (dashed blue line) and $\eta_{i,\theta} > 0$ for some $\theta \in [0,1]$ (solid blue line). In general equilibrium, migration lowers optimal marginal tax rates only for some individuals. Both tax schedules are U-shaped up to $\$200\,000$, reflecting the shape of the hazard rate of the income distribution. As

high-skilled immigrants in Denmark, find substantially larger migration elasticities.

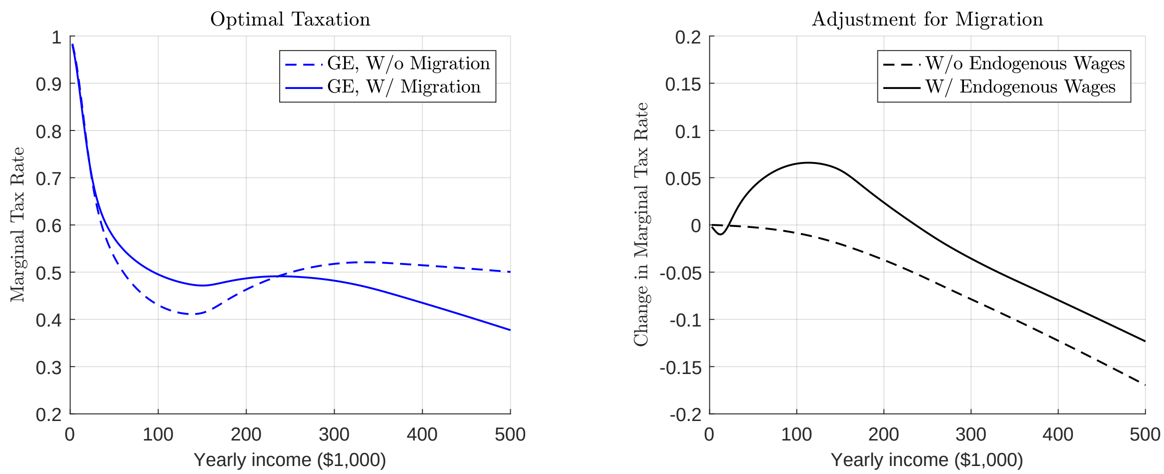


Figure 2: Left Panel: Optimal Taxation in General Equilibrium W/ and W/o Migration; Right Panel: Adjustment in Optimal Taxation for Migration in PE and GE; Increasing Migration Semi-Elasticity

in Lehmann et al. (2014), migration causes the optimal marginal tax rates to drop rapidly for top earners. Relative to the case without migration, the majority of workers (up to around $\$230\,000$) experiences a rise in tax rates (unlike in Lehmann et al. (2014)). At the top, workers face lower tax rates (in line with Lehmann et al. (2014)). However, at the very top (above $\$1\,600\,000$, not shown), the marginal tax rates are again higher with migration than without migration, which is the result already found in Proposition 1. The main takeaway of this simulation is that migration may increase tax rates for large parts of the income distribution.

The solid line in the right panel of Figure 2 computes the migration-induced adjustment in tax rates in general equilibrium. The dashed line depicts the same adjustment but in a partial equilibrium setting. As described, there is a positive adjustment in general equilibrium, where the partial equilibrium economy would predict a decline. Also, in those parts where the signs of the adjustments are aligned the migration-induced reduction in tax rates is less pronounced in general than in partial equilibrium. The transfer level in general equilibrium declines by only $\$95$, compared to a more pronounced reduction of $\$283$ in partial equilibrium. Altogether, the presence of general equilibrium effects makes the migration threat less severe.

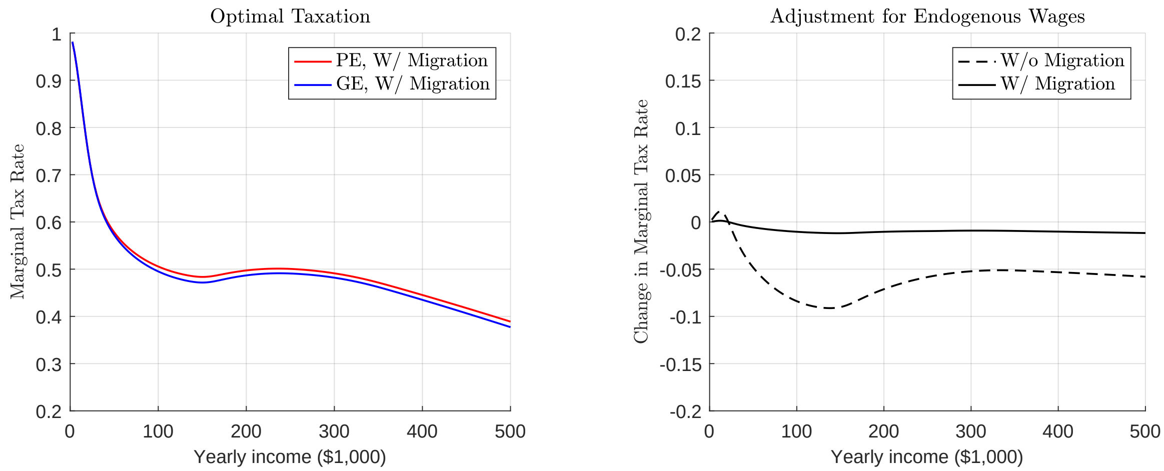


Figure 3: Left Panel: Optimal Taxation under Migration W/ and W/o Endogenous Wages; Right Panel: Adjustment in Optimal Taxation for Endogenous Wages W/ and W/o Migration; Increasing Migration Semi-Elasticity

4.3 Presence of Endogenous Wages

We may also study how the presence of general equilibrium responses affects the optimal tax scheme. In the left panel of Figure 3, we show optimal tax rates with migration in partial equilibrium (red line, $\sigma = \infty$) and in general equilibrium (blue line, $\sigma = 2$). Unlike in Sachs et al. (2020), general equilibrium effects have very little effect on the optimal tax scheme. If at all, there is a miniscule downward adjustment for high incomes. This finding also appears in the right panel of the figure (solid line). The dashed line contrasts this with the closed-economy setting, where the usual decline in tax rates appears (see Sachs et al. (2020)). We conclude that migration blunts the general equilibrium effects and, therefore, the trickle-down rationale.

4.4 Effect Decomposition

In Figure 4, we decompose the revenue impact at the optimal tax scheme in general equilibrium with migration into each of the effects described in our incidence analysis. Again, we consider an elementary tax reform that uniformly raises tax payments above a certain income level (x-axis) and evaluate it at the optimal tax scheme with migration and general equilibrium effects. As before, we normalize the reform's mechanical effect on government revenues to 1\$. On the y-axis, we depict how much each of the effects adds to government revenues (in \$). By

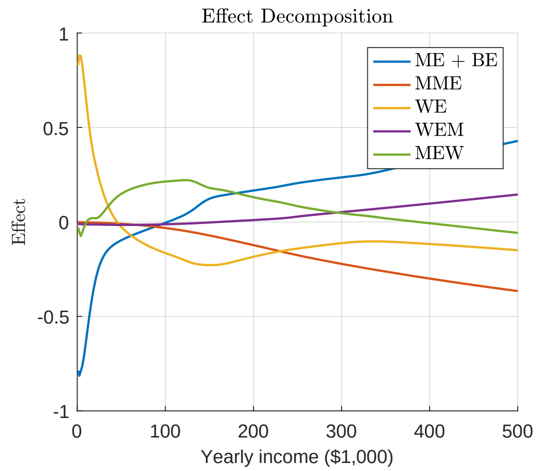


Figure 4: Effect Decomposition: Mechanical and Behavioral Effect (ME + BE), Mechanical Migration Effect (MME), Wage Effect (WE), Wage Effect on Migration (WEM), Migration Effect on Wages (MEW); Increasing Migration Semi-Elasticity

construction of the optimal tax scheme, the effects add up to zero. One can interpret the figure as follows. If there were no migration or general equilibrium effects (no predistribution), the combined mechanical and behavioral effect (blue line, see Saez (2001)) would call for lower tax rates at the bottom and higher ones at the top. The mechanical migration effect (orange line, see Lehmann et al. (2014)) pushes tax rates down starting at $\$100\,000$, whereas the wage effect (yellow line, see Sachs et al. (2020)) calls for higher taxes at the bottom and lower ones at the top. The novel migration effect on wages (green line) is particularly important for the upper middle class and the wage effect on migration (purple line) pushes taxes up in higher parts of the income distribution.

5 Extensions

Tax coordination. Our setup allows us to consider the effects of tax coordination on marginal income tax rates. Tax coordination provides a way to overcome the inefficiencies from the non-cooperative setting of tax policies, although typically in the context of representative household models. In Appendix A.3, we show that in our discrete-type framework of Section 2, under cross-country symmetry, governments can restore the autarky solution by coordinating their income taxation. The intuition is that governments internalize the cross-country externalities from international labor migration when coordinating their tax policies. Thus, in general equilibrium, international coordination of income taxation leads to *less* tax progressivity in terms of marginal tax rates. This finding is in contrast to the conventional view that fiscal competition between governments limits the amount of redistribution, and tax coordination may, therefore, raise the level of tax progressivity (for a survey of the literature on tax competition and coordination, see Keen and Konrad (2013)). Moreover, notice that the two-type version of the coordinated tax policy setup is equivalent to Rothschild and Scheuer (2013). In their model, a policymaker sets a tax scheme under occupational mobility. In our coordination setting, a planner chooses the tax system in both countries subject to the international mobility of labor.

Alternative welfare criteria. First, note that our tax incidence analysis in Section 3.2 is free of assumptions on social welfare. However, to derive the optimal tax formulas in Section 3.3, we need to take a stand on the government's objective function. The choice of the social welfare function is, in the context of migration, non-trivial because one has to take a stand on how the government treats the utility, in particular, of those workers who are not present in a country at the moment the policies are chosen.[26] There are, at least, three critical dimensions (for an overview, see, for instance, Simula and Trannoy (2012)). Firstly, one needs to decide how much weight to put on each type making interpersonal comparisons. Our Rawlsian objective puts all weight on the worst-off type. A second question is whether to maximize the welfare of citizens or residents. Our Rawlsian criterion abstracts from this issue by implicitly assuming that the lowest type in the income distribution is always present. Thirdly, one has to specify whether workers' average or total utility enters social welfare. We decided for the average utility (of the lowest type). In Appendix B.5, we depart in the continuous-type model from our assumptions in each of these dimensions by considering the aggregate welfare $\mathcal{G}_i \equiv \int_{\theta \in \Theta} \Gamma_i \left(\mathcal{U}_{i,\theta}\right) N_{i,\theta} d\theta$, where $\Gamma_i\left(\cdot\right)$ is an increasing and concave generalized social welfare function (see Saez and Stantcheva (2016)). We derive the optimal taxation when the government maximizes the weighted sum of residents' utilities. While the analysis yields similar results, we admit that the choice of the welfare criterion remains arbitrary and may lead to different results.

[26] A similar issue is discussed in the literature on endogenous fertility (see, e.g., Golosov, Jones, and Tertilt (2007)).

6 Conclusion

In this paper, we introduce migration into a model of nonlinear taxation in general equilibrium. By adding an extensive margin to our K-type model and the continuous-type framework, we make the canonical Stiglitz (1982) and Mirrlees (1971) models with endogenous labor supply and wages more realistic. As we have shown, contrary to conventional wisdom, migration may lead to a more progressive tax code in terms of marginal tax rates. This finding is at odds with the notion that a migration threat reduces marginal tax rates. Moreover, we show that migration responses mitigate the trickle-down rationale in general equilibrium. Thus, even though migration and general equilibrium forces are, in isolation, known to limit the degree of redistribution, when considering them jointly, they are partly offsetting each other.

Our result suggests an alternative explanation of the recent empirical literature on the effects of globalization on redistribution and inequality. Most prominently, Egger, Nigai, and Strecker (2019) demonstrate that increases in both international trade and migration in OECD countries in the 1980s and early 1990s led to higher average tax burdens (but less in the following years). A well-known explanation for this finding is that redistribution, as well as government size (e.g., Rodrik (1998)), compensates for the adverse effects of globalization on workers from lower parts of the income distribution (see Autor, Dorn, and Hanson (2013), for study on the labor market effects, and Schulz, Tsyvinski, and Werquin (2022), for a solution to the compensation problem in general equilibrium). Our finding does not call into question this widespread view. Instead, it offers an alternative explanation for why globalization may lead to higher tax rates along with a rise in wage inequality. The main difference is that, in the former view, globalization directly amplifies pre-existing inequities, whereas, in our framework, the policy response to rising international mobility induces greater inequality.

References

Ales, L., Kurnaz, M., and Sleet, C. (2015). "Technical change, wage inequality, and taxes." *American Economic Review*, *105*(10), 3061–3101.

Autor, D., Dorn, D., and Hanson, G. H. (2013). "The china syndrome: Local labor market effects of import competition in the united states." *American Economic Review*, *103*(6), 2121–68.

Benabou, R. (2000). "Unequal societies: Income distribution and the social contract." *American Economic Review*, *90*(1), 96–129.

Bierbrauer, F., Brett, C., and Weymark, J. A. (2013). "Strategic nonlinear income tax competition with perfect labor mobility." *Games and Economic Behavior*, *82*, 292–311.

Bound, J., and Holzer, H. J. (2000). "Demand shifts, population adjustments, and labor market outcomes during the 1980s." *Journal of Labor Economics*, *18*(1), 20–54.

Card, D. (2009). "Immigration and inequality." *American Economic Review*, *99*(2), 1–21.

Card, D., and Lemieux, T. (2001). "Can falling supply explain the rising return to college for younger men? a cohort-based analysis." *The Quarterly Journal of Economics*, *116*(2), 705–746.

Chetty, R. (2012). "Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply." *Econometrica*, *80*(3), 969–1018.

Colas, M., and Sachs, D. (2020). "The indirect fiscal benefits of low-skilled immigration." *CESifo Working Paper No. 8604*.

Diamond, P. A. (1998). "Optimal income taxation: an example with a u-shaped pattern of optimal marginal tax rates." *American Economic Review*, 83–95.

Diamond, R. (2016). "The determinants and welfare implications of us workers' diverging location choices by skill: 1980-2000." *American Economic Review*, *106*(3), 479–524.

Dustmann, C., Frattini, T., and Preston, I. P. (2013). "The effect of immigration along the distribution of wages." *Review of Economic Studies*, *80*(1), 145–173.

Egger, P. H., Nigai, S., and Strecker, N. M. (2019). "The taxing deed of globalization." *American Economic Review*, *109*(2), 353–90.

Feldstein, M. S. (1969). "The effects of taxation on risk taking." *Journal of Political Economy*, *77*(5), 755–764.

Glazer, A., Kanniainen, V., and Poutvaara, P. (2008). "Income taxes, property values, and migration." *Journal of Public Economics*, *92*(3-4), 915–923.

Golosov, M., Jones, L., and Tertilt, M. (2007). "Efficiency with endogenous population growth." *Econometrica*, *75*, 1039–1071.

Greenwood, J., Hercowitz, Z., and Huffman, G. W. (1988). "Investment, capacity utilization, and the real business cycle." *American Economic Review*, 402–417.

Guerreiro, J., Rebelo, S., and Teles, P. (2020). "What is the optimal immigration policy? migration, jobs, and welfare." *Journal of Monetary Economics*, *113*, 61–87.

Heathcote, J., Storesletten, K., and Violante, G. L. (2017). "Optimal tax progressivity: An analytical framework." *The Quarterly Journal of Economics*, *132*(4), 1693–1754.

Heathcote, J., and Tsujiyama, H. (2021a). "Optimal income taxation: Mirrlees meets ramsey." *Journal of Political Economy*, *129*(11), 3141–3184.

Heathcote, J., and Tsujiyama, H. (2021b). "Practical optimal income taxation." *Working Paper*.

Jacquet, L., and Lehmann, E. (2021). "Optimal income taxation with composition effects." *Journal of the European Economic Association*, *19*(2), 1299–1341.

Janeba, E., and Schulz, K. (2021). "Nonlinear taxation and international mobility in general equilibrium." *Working Paper*.

Katz, L. F., and Murphy, K. M. (1992). "Changes in relative wages, 1963–1987: supply and demand factors." *The Quarterly Journal of Economics*, *107*(1), 35–78.

Keen, M., and Konrad, K. A. (2013). "The theory of international tax competition and coordination." *Handbook of Public Economics*, *5*, 257–328.

Kleven, H., Landais, C., Munoz, M., and Stantcheva, S. (2020). "Taxation and migration: Evidence and policy implications." *Journal of Economic Perspectives*, *34*(2), 119–42.

Kleven, H. J., Landais, C., Saez, E., and Schultz, E. (2014). "Migration and wage effects of taxing top earners: Evidence from the foreigners' tax scheme in denmark." *The Quarterly Journal of Economics*, *129*(1), 333–378.

Lehmann, E., Simula, L., and Trannoy, A. (2014). "Tax me if you can! optimal nonlinear income tax between competing governments." *The Quarterly Journal of Economics*, *129*(4), 1995–2030.

Lessem, R. (2018). "Mexico–us immigration: effects of wages and border enforcement." *The Review of Economic Studies*, *85*(4), 2353–2388.

Milgrom, P., and Shannon, C. (1994). "Monotone comparative statics." *Econometrica*, 157–180.

Mirrlees, J. A. (1971). "An exploration in the theory of optimum income taxation." *The Review of Economic Studies*, *38*(2), 175–208.

Notowidigdo, M. J. (2020). "The incidence of local labor demand shocks." *Journal of Labor Economics*, *38*(3), 687–725.

Persson, M. (1983). "The distribution of abilities and the progressive income tax." *Journal of Public Economics*, *22*(1), 73–88.

Piyapromdee, S. (2021). "The impact of immigration on wages, internal migration, and welfare." *The Review of Economic Studies*, *88*(1), 406–453.

Rodrik, D. (1998). "Why do more open economies have bigger governments?" *Journal of Political Economy*, *106*(5), 997–1032.

Rothschild, C., and Scheuer, F. (2013). "Redistributive taxation in the roy model." *The Quarterly Journal of Economics*, *128*(2), 623–668.

Rothschild, C., and Scheuer, F. (2016). "Optimal taxation with rent-seeking." *The Review of Economic Studies*, *83*(3), 1225–1262.

Sachs, D., Tsyvinski, A., and Werquin, N. (2020). "Nonlinear tax incidence and optimal taxation in general equilibrium." *Econometrica*, *88*(2), 469–493.

Saez, E. (2001). "Using elasticities to derive optimal income tax rates." *The Review of Economic Studies*, *68*(1), 205–229.

Saez, E., and Stantcheva, S. (2016). "Generalized social marginal welfare weights for optimal tax theory." *American Economic Review*, *106*(1), 24–45.

Scheuer, F., and Werning, I. (2017). "The taxation of superstars." *The Quarterly Journal of Economics*, *132*(1), 211–270.

Schulz, K., Tsyvinski, A., and Werquin, N. (2022). "Generalized compensation principle." *FRB of Chicago Working Paper*.

Simula, L., and Trannoy, A. (2010). "Optimal income tax under the threat of migration by top-income earners." *Journal of Public Economics*, *94*(1-2), 163–173.

Simula, L., and Trannoy, A. (2012). "Shall we keep the highly skilled at home? the optimal income tax perspective." *Social Choice and Welfare*, *39*, 751–782.

Stiglitz, J. E. (1982). "Self-selection and pareto efficient taxation." *Journal of Public Economics*, *17*(2), 213–240.

Tsugawa, S. (2021). "Income tax competition with endogenous wages." *Working Paper*.

A Proofs for the *K*-Type Model

A.1 Proof of Proposition 1

The optimal tax code can be implicitly described by the households' first-order condition (2). In the following, we employ a decomposition into an "inner" and an "outer" problem. Firstly, we characterize the solution to the "inner" problem. That is, we solve for the optimal allocation $\{c_{i,\theta}, l_{i,\theta}\}_{\theta\in\Theta}$ for a given population $\{N_{i,\theta}\}_{\theta\in\Theta}$. Secondly, we maximize welfare by choosing $\{N_{i,\theta}\}_{\theta\in\Theta}$, which is the "outer" problem.

Inner problem. The Lagrangian function of the benevolent social planner in country i is defined by

$$\begin{aligned}\mathcal{L}_i\left(\{N_{i,\theta}\}_{\theta\in\Theta}\right) \equiv & u\left(c_{i,1}, l_{i,1}\right) + \sum_{\theta\in\{2,\dots,K\}} \mu_{i,\theta}\left[u\left(c_{i,\theta}, l_{i,\theta}\right) - u\left(c_{i,\theta-1}, \frac{l_{i,\theta-1} w_{i,\theta-1}}{w_{i,\theta}}\right)\right] \\ & + \xi_i\left[F_i\left(\{N_{i,\theta} l_{i,\theta}\}_{\theta\in\Theta}\right) - \sum_{\theta\in\Theta} N_{i,\theta} c_{i,\theta}\right] + \sum_{\theta\in\Theta} \lambda_{i,\theta}\left[N_{i,\theta} - \rho_i\left(\Delta_{i,\theta};\theta\right)\right]\end{aligned}$$

for a given population. Assuming that the optimization problem is convex and using the definitions of wages, wage elasticities, and migration semi-elasticities, the following first-order conditions describe the unique optimum

$$\begin{aligned}[c_{i,\theta}] : 0 = & \mathbb{1}\left[\theta = 1\right] u_c\left(c_{i,\theta}, l_{i,\theta}\right) - N_{i,\theta}\xi_i + \mathbb{1}\left[\theta > 1\right]\mu_{i,\theta} u_c\left(c_{i,\theta}, l_{i,\theta}\right) - \mathbb{1}\left[\theta < K\right]\mu_{i,\theta+1} u_c\left(c_{i,\theta}, \frac{y_{i,\theta}}{w_{i,\theta+1}}\right) \\ & - \lambda_{i,\theta}\eta_{i,\theta} N_{i,\theta} u_c\left(c_{i,\theta}, l_{i,\theta}\right) \qquad (21)\end{aligned}$$

$$\begin{aligned}[l_{i,\theta}] : 0 = & \mathbb{1}\left[\theta = 1\right] u_l\left(c_{i,\theta}, l_{i,\theta}\right) + \mathbb{1}\left[\theta > 1\right]\mu_{i,\theta} u_l\left(c_{i,\theta}, l_{i,\theta}\right) - \mathbb{1}\left[\theta < K\right]\mu_{i,\theta+1} u_l\left(c_{i,\theta}, \frac{y_{i,\theta}}{w_{i,\theta+1}}\right)\frac{w_{i,\theta}}{w_{i,\theta+1}} \\ & + \xi_i w_{i,\theta} N_{i,\theta} - \lambda_{i,\theta}\eta_{i,\theta} N_{i,\theta} u_l\left(c_{i,\theta}, l_{i,\theta}\right) \\ & - \sum_{k\in\{2,\dots,K\}} \mu_{i,k} u_l\left(c_{i,k-1}, \frac{y_{i,k-1}}{w_{i,k}}\right)\frac{y_{i,k-1}}{l_{i,\theta} w_{i,k}}\left(\gamma_{i,k-1,\theta} - \gamma_{i,k,\theta}\right), \qquad (22)\end{aligned}$$

for $\theta \in \Theta$, where $\mathbb{1}\left[\cdot\right]$ is the indicator function. Inserting Equation (21) into (22) and making use of the worker's first-order condition, the marginal tax rate of a worker θ can be written as

$$T_i'(y_{i,\theta}) = \frac{\mathbb{1}[\theta < K] \frac{\mu_{i,\theta+1}}{\xi_i} \frac{u_c\left(c_{i,\theta}, \frac{l_{i,\theta} w_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}} \left[1 + \frac{u_l\left(c_{i,\theta}, \frac{l_{i,\theta} w_{i,\theta}}{w_{i,\theta+1}}\right)}{w_{i,\theta+1} u_c\left(c_{i,\theta}, l_{i,\theta}\right)}\right]}{1 + \mathbb{1}[\theta < K] \frac{\mu_{i,\theta+1}}{\xi_i} \frac{u_c\left(c_{i,\theta}, \frac{l_{i,\theta} w_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}}$$

$$+ \frac{\frac{1}{y_{i,\theta} N_{i,\theta}} \sum_{k \in \{2,\ldots,K\}} \frac{\mu_{i,k}}{\xi_i} u_l\left(c_{i,k-1}, \frac{l_{i,k-1} w_{i,k-1}}{w_{i,k}}\right) \frac{y_{i,k-1}}{w_{i,k}} \left(\gamma_{i,k-1,\theta} - \gamma_{i,k,\theta}\right)}{1 + \mathbb{1}[\theta < K] \frac{\mu_{i,\theta+1}}{\xi_i} \frac{u_c\left(c_{i,\theta}, \frac{l_{i,\theta} w_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}}.$$

In their Proposition (1), Ales et al. (2015) show that one can decompose the formula for the optimal tax rate in the K-type Stiglitz (1982) setting without migration ($\eta_{i,\theta} = 0, \ \forall \theta \in \Theta$) into a Mirrleesian and a wage compression term. A similar decomposition applies here with the difference that one needs to account for migration responses ($\eta_{i,\theta} > 0$), affecting the Lagrange multipliers $\frac{\mu_{i,\theta}}{\xi_i}$ and, indirectly, the equilibrium population $N_{i,\theta}$.

On the one hand, the Mirrleesian term is augmented by the direct partial equilibrium impact of migration on the objective function, which Lehmann et al. (2014) label as the "migration threat." On the other hand, migration interacts with the wage compression term. Thus, this setting nests the well-known partial equilibrium effect of migration on the optimal taxation and adds general equilibrium moderation effects. In the following, we derive conditions under which the classical partial equilibrium downward force of migration on taxes is offset by our novel general equilibrium moderation effects.

Under a CES production function, this expression for the optimal marginal tax rate simplifies to

$$T_i'(y_{i,\theta}) = \frac{\mathbb{1}[\theta < K] \frac{\mu_{i,\theta+1}}{\xi_i} \frac{u_c\left(c_{i,\theta}, \frac{y_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}} \left[1 + \frac{u_l\left(c_{i,\theta}, \frac{y_{i,\theta}}{w_{i,\theta+1}}\right)}{w_{i,\theta+1} u_c\left(c_{i,\theta}, l_{i,\theta}\right)} \frac{\sigma - 1}{\sigma}\right] + \mathbb{1}[\theta > 1] \frac{\mu_{i,\theta}}{\xi_i} \frac{y_{i,\theta-1}}{y_{i,\theta} N_{i,\theta}} \frac{1}{\sigma} \frac{u_l\left(c_{i,\theta-1}, \frac{y_{i,\theta-1}}{w_{i,\theta}}\right)}{w_{i,\theta}}}{1 + \mathbb{1}[\theta < K] \frac{\mu_{i,\theta+1}}{\xi_i} \frac{u_c\left(c_{i,\theta}, \frac{y_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}}.$$

Therefore, setting $\theta = 1$ and $\theta = K$, the bottom and top tax rates are characterized by

$$T_i'(y_{i,1}) = \frac{\frac{\mu_{i,2}}{\xi_i} \frac{u_c\left(c_{i,1}, \frac{y_{i,1}}{w_{i,2}}\right)}{N_{i,1}}}{1 + \frac{\mu_{i,2}}{\xi_i} \frac{u_c\left(c_{i,1}, \frac{y_{i,1}}{w_{i,2}}\right)}{N_{i,1}}} \left[1 + \frac{u_l\left(c_{i,1}, \frac{y_{i,1}}{w_{i,2}}\right)}{w_{i,2} u_c\left(c_{i,1}, l_{i,1}\right)} \frac{\sigma - 1}{\sigma}\right]$$

and

$$T_i'\left(y_{i,K}\right)=\frac{\mu_{i,K}}{\xi_i}\frac{y_{i,K-1}}{y_{i,K}N_{i,K}}\frac{1}{\sigma}\frac{u_l\left(c_{i,K-1},\frac{y_{i,K-1}}{w_{i,K}}\right)}{w_{i,K}}$$

respectively. These expressions depend on the (relative) Lagrange multipliers $\frac{\mu_{i,\theta}}{\xi_i}$.

One can obtain the shadow value of public funds by summing up Equation (21) over all types and using the separability of consumption and leisure, which yields $\xi_i=\frac{1-\sum_{\theta\in\Theta}\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}}{\sum_{\theta\in\Theta}N_{i,\theta}/u_c\left(c_{i,\theta},l_{i,\theta}\right)}$. Plugging this value back into (21), the normalized multiplier on the incentive constraint for $k\in\{2,...,K\}$ reads as

$$\frac{\mu_{i,k}}{\xi_i}=\frac{\sum_{l=k}^K N_{i,l}/u_c\left(c_{i,l},l_{i,l}\right)}{1-\sum_{\theta\in\Theta}\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}}\left(1+\frac{\sum_{l=1}^{k-1}N_{i,l}/u_c\left(c_{i,l},l_{i,l}\right)}{\sum_{l=k}^{K}N_{i,l}/u_c\left(c_{i,l},l_{i,l}\right)}\sum_{l=k}^{K}\lambda_{i,l}\eta_{i,l}N_{i,l}-\sum_{l=1}^{k-1}\lambda_{i,l}\eta_{i,l}N_{i,l}\right)\equiv\frac{\mu_{i,k}^{ex}}{\xi_i^{ex}}\mathcal{S}_{i,k},$$

where we define the normalized multiplier without migration responses ($\eta_{i,\theta}=0,\ \forall\theta\in\Theta$) as

$$\frac{\mu_{i,k}^{ex}}{\xi_i^{ex}}\equiv\sum_{l=k}^{K}N_{i,l}/u_c\left(c_{i,l},l_{i,l}\right)$$

and a scaling factor as

$$\mathcal{S}_{i,k}\equiv\frac{1+\frac{\sum_{l=1}^{k-1}N_{i,l}/u_c\left(c_{i,l},l_{i,l}\right)}{\sum_{l=k}^{K}N_{i,l}/u_c\left(c_{i,l},l_{i,l}\right)}\sum_{l=k}^{K}\lambda_{i,l}\eta_{i,l}N_{i,l}-\sum_{l=1}^{k-1}\lambda_{i,l}\eta_{i,l}N_{i,l}}{1-\sum_{\theta\in\Theta}\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}}.$$

This procedure gives us the exogenous technology planner's optimal bottom and top tax rates

$$T_i^{ex\prime}\left(y_{i,1}\right)\equiv\frac{\frac{\mu_{i,2}^{ex}}{\xi_i^{ex}}\frac{u_c\left(c_{i,1},\frac{y_{i,1}}{w_{i,2}}\right)}{N_{i,1}}}{1+\frac{\mu_{i,2}^{ex}}{\xi_i^{ex}}\frac{u_c\left(c_{i,1},\frac{y_{i,1}}{w_{i,2}}\right)}{N_{i,1}}}\left[1+\frac{u_l\left(c_{i,1},\frac{y_{i,1}}{w_{i,2}}\right)}{w_{i,2}u_c\left(c_{i,1},l_{i,1}\right)}\frac{\sigma-1}{\sigma}\right]>0$$

and

$$T_i^{ex\prime}\left(y_{i,K}\right)\equiv\frac{\mu_{i,K}^{ex}}{\xi_i^{ex}}\frac{y_{i,K-1}}{y_{i,K}N_{i,K}}\frac{1}{\sigma}\frac{u_l\left(c_{i,K-1},\frac{y_{i,K-1}}{w_{i,K}}\right)}{w_{i,K}}<0,$$

to which we can now compare the Nash equilibrium tax rates. The comparison between $T_i'\left(y_{i,\theta}\right)$ and $T_i^{ex\prime}\left(y_{i,\theta}\right)$ depends on the adjustment in the Lagrange multipliers measured by the scaling factor $\mathcal{S}_{i,k}$ (i.e., $\frac{\mu_{i,\theta}}{\xi_i}$ vs. $\frac{\mu_{i,\theta}^{ex}}{\xi_i^{ex}}$).[27] More precisely, for Propositions 1, we need to show that $\mathcal{S}_{i,K}<$

[27]Notice that the right-hand side of $T_i^{ex\prime}\left(y_{i,\theta}\right)$ depends on the equilibrium allocation that may be endogenous to tax policy. For simplicity, we evaluate, in the following comparison, the right-hand side at the allocation chosen in the Nash equilibrium. In Appendix A.2, we provide conditions under which the right-hand side is given in closed

1 (part a)) and $\mathcal{S}_{i,2} < 1$ (part b)), which holds for $\lambda_{i,K}\eta_{i,K}N_{i,K} < 0$ and $\sum_{l=2}^{K} \lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta} < 0$, respectively.

Outer problem. To prove these statements, we now derive the first-order conditions with respect to the population masses

$$
\begin{aligned}
[N_{i,\theta}] : 0 = &- \sum_{k\in\{2,\ldots,K\}} \mu_{i,k} u_l \left(c_{i,\theta-1}, \frac{l_{i,k-1}w_{i,k-1}}{w_{i,k}}\right) \frac{l_{i,k-1}w_{i,k-1}}{N_{i,\theta}w_{i,k}} \left(\gamma_{i,k-1,\theta} - \gamma_{i,k,\theta}\right) \\
&+ \xi_i \left(l_{i,\theta}w_{i,\theta} - c_{i,\theta}\right) + \lambda_{i,\theta}.
\end{aligned} \tag{23}
$$

For a constant elasticity of substitution production function, Equation (23) simplifies to

$$
-\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta} = \xi_i T_i\left(y_{i,\theta}\right) N_{i,\theta}\eta_{i,\theta} + \frac{1}{\sigma}\eta_{i,\theta}\left[\mu_{i,\theta+1}u_l\left(c_{i,\theta}, \frac{y_{i,\theta}}{w_{i,\theta+1}}\right)\frac{y_{i,\theta}}{w_{i,\theta+1}} - \mu_{i,\theta}u_l\left(c_{i,\theta-1}, \frac{y_{i,\theta-1}}{w_{i,\theta}}\right)\frac{y_{i,\theta-1}}{w_{i,\theta}}\right].
$$

Noting that $T_i\left(y_{i,K}\right) \geq 0$, since $T_i\left(y_{i,\theta}\right) \geq T_i\left(y_{i,\theta-1}\right)$ and $\sum_{\theta\in\Theta} T_i\left(y_{i,\theta}\right) N_{i,\theta} \geq 0$, we conclude that

$$
-\lambda_{i,K}\eta_{i,K}N_{i,K} = \xi_i T_i\left(y_{i,K}\right) N_{i,K}\eta_{i,K} - \frac{1}{\sigma}\eta_{i,K}\mu_{i,K}u_l\left(c_{i,K-1}, \frac{y_{i,K-1}}{w_{i,K}}\right)\frac{y_{i,K-1}}{w_{i,K}} > 0,
$$

or $\mathcal{S}_{i,K} < 1$ (part a)). Thus, we do not rely on any assumption about the migration semi-elasticity for the result about the top tax rate (first part of Proposition 1).

As mentioned above, for the decline in bottom tax rate postulated in the second part of Proposition 1, we need to show that $\mathcal{S}_{i,2} < 1$ or

$$
-\sum_{\theta=2}^{K} \lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta} = -\sum_{\theta=1}^{K} \lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta} + \lambda_{i,1}\eta_{i,1}N_{i,1} > 0.
$$

Let the assumption stated in Proposition 1 hold. Then, in the symmetric Nash equilibrium

$$
\begin{aligned}
-\sum_{\theta=1}^{K} \lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta} = &\ \xi_i \sum_{\theta=1}^{K} T_i\left(y_{i,\theta}\right) n_{i,\theta}\eta_{i,\theta} \\
&+ \frac{1}{\sigma}\sum_{\theta=1}^{K}\eta_{i,\theta}\left[\mu_{i,\theta+1}u_l\left(c_{i,\theta}, \frac{y_{i,\theta}}{w_{i,\theta+1}}\right)\frac{y_{i,\theta}}{w_{i,\theta+1}} - \mu_{i,\theta}u_l\left(c_{i,\theta-1}, \frac{y_{i,\theta-1}}{w_{i,\theta}}\right)\frac{y_{i,\theta-1}}{w_{i,\theta}}\right]
\end{aligned}
$$

The second summand is positive for $\eta_{i,\theta} \geq \eta_{i,\theta-1}$. For a constant population size ($n_{i,\theta} = n_{i,\theta-1}$),

form and, thus, independent from the equilibrium allocation.

the first summand can be written as

$$\xi_i n_{i,\theta} \sum_{\theta=1}^{K} T_i(y_{i,\theta}) \eta_{i,\theta} \geq \left(\frac{1}{K} \sum_{\theta=1}^{K} \eta_{i,\theta} \right) \xi_i \sum_{\theta=1}^{K} T_i(y_{i,\theta}) n_{i,\theta} = 0$$

where the first inequality is Chebyshev's sum inequality for $\eta_{i,\theta} \geq \eta_{i,\theta-1}$ and $T_i(y_{i,\theta}) \geq T_i(y_{i,\theta-1})$ and the second inequality follows from the government's budget constraint. Therefore, $-\sum_{\theta=1}^{K} \lambda_{i,\theta} \eta_{i,\theta} N_{i,\theta} \geq 0$. To conclude that $-\sum_{\theta=2}^{K} \lambda_{i,\theta} \eta_{i,\theta} N_{i,\theta} > 0$, notice that

$$\lambda_{i,1} \eta_{i,1} N_{i,1} = -\xi_i T_i(y_{i,1}) n_{i,1} \eta_{i,1} - \frac{1}{\sigma} \eta_{i,1} \mu_{i,2} u_l \left(c_{i,1}, \frac{y_{i,1}}{w_{i,2}} \right) \frac{y_{i,1}}{w_{i,2}} > 0$$

since $T_i(y_{i,1}) \leq 0$, for $\sum_{\theta=1}^{K} T_i(y_{i,\theta}) N_{i,\theta} \geq 0$ and $T_i(y_{i,\theta}) \geq T_i(y_{i,\theta-1})$, and $u_l \left(c_{i,1}, \frac{y_{i,1}}{w_{i,2}} \right) < 0$. Therefore, $\mathcal{S}_{i,2} < 1$ and part $b)$ of the proposition holds.

A.2 A Closed-Form Example

The purpose of this section is to provide an example in which one obtains closed-form expressions for the optimal tax rates chosen by the exogenous technology planner. Let $K = 2$ and label L and H as the low- and high-skill worker type. Suppose that $F_i(N_{i,L} l_{i,L}, N_{i,H} l_{i,H}) = A_i (N_{i,L} l_{i,L})^{\alpha} (N_{i,H} l_{i,H})^{1-\alpha}$ for $\alpha \in (0,1)$. Then, the own- and cross-wage elasticities are given by $\gamma_{i,L,L} = -(1-\alpha)$, $\gamma_{i,H,H} = -\alpha$, $\gamma_{i,L,H} = 1-\alpha$, and $\gamma_{i,H,L} = \alpha$. The income share of the unskilled relative to the skilled workers reads as $\frac{N_{i,L} l_{i,L} w_{i,L}}{N_{i,H} l_{i,H} w_{i,H}} = \frac{\alpha}{1-\alpha}$. Moreover, let the consumption utility be linear $u(c) = c$, and assume an isoelastic disutility from labor $v(l) = \frac{l^{1+1/e}}{1+1/e}$ with e denoting the Frisch elasticity of labor supply. Finally, consider a setup with symmetric countries—thus, the symmetric Nash equilibrium in which no mobility occurs ($N_{i,\theta} = n_{i,\theta}$ for $i = A, B$ and $\theta = L, H$).

Then, one can write the exogenous technology planner's marginal tax rate for the high-skilled workers as $T_i^{ex\prime}(y_{i,H}) = -\left(\frac{\alpha}{1-\alpha} \frac{n_{i,H}}{n_{i,L}} \right)^{1+1/e} \frac{l_{i,H}^{1/e}}{w_{i,H}}$. Using the workers' first-order condition, the marginal tax rate at the top simplifies to

$$\frac{T_i^{ex\prime}(y_{i,H})}{1 - T_i^{ex\prime}(y_{i,H})} = -\left(\frac{\alpha}{1-\alpha} \frac{n_{i,H}}{n_{i,L}} \right)^{1+1/e} < 0.$$

Applying similar steps, the marginal tax rate for low-skilled workers reads as

$$\frac{T_i^{ex\prime}(y_{i,L})}{1 - T_i^{ex\prime}(y_{i,L})} = \frac{n_{i,H}}{n_{i,L}} > 0.$$

Now, we show that, in this parametrization, there is a negative reduced-form relationship between high-skilled workers' gross income and their marginal tax rate. Notice that this exercise is non-trivial in our setup since one needs to consider general equilibrium wage effects from labor supply and migration. By a high-skilled worker's first-order condition $(l_{i,H})^{1/e} = w_{i,H}\left[1 - T'(y_{i,H})\right]$, the income response to a cut in the top tax rate depends on a direct labor supply and an indirect wage response $\frac{d\ln\left(y_{i,H}\right)}{d\ln\left[1-T'\left(y_{i,H}\right)\right]} = (1+e)\frac{d\ln\left(w_{i,H}\right)}{d\ln\left[1-T'\left(y_{i,H}\right)\right]} + e$, whose overall sign is not clear a priori. To calculate the indirect wage response, we first derive a low-skilled worker's consumption, income, and wage response

$$\frac{d\,c_{i,L}}{d\left[1 - T'(y_{i,H})\right]} = \frac{1 - T'(y_{i,L})}{1 - T'(y_{i,H})} y_{i,L}(1+e)\frac{d\ln(w_{i,L})}{d\ln\left[1 - T'(y_{i,H})\right]},$$

$$\frac{d\,l_{i,L}}{d\left[1 - T'(y_{i,H})\right]} = \frac{1 - T'(y_{i,L})}{1 - T'(y_{i,H})} l_{i,L} e\frac{d\ln(w_{i,L})}{d\ln\left[1 - T'(y_{i,H})\right]},$$

and

$$\frac{d\ln(w_{i,L})}{d\ln\left[1 - T'(y_{i,H})\right]} = \frac{d\ln\left[\alpha A_i\left(N_{i,L}l_{i,L}\right)^{\alpha-1}\left(N_{i,H}l_{i,H}\right)^{1-\alpha}\right]}{d\ln\left[1 - T'(y_{i,H})\right]} = -\frac{1-\alpha}{\alpha}\frac{d\ln(w_{i,H})}{d\ln\left[1 - T'(y_{i,H})\right]}.$$

Accordingly, a high-skilled worker's wages change as follows

$$\begin{aligned}\frac{d\ln(w_{i,H})}{d\ln\left[1 - T'(y_{i,H})\right]} = {} & \frac{\alpha}{1+\alpha e}\eta_{i,L}\left[1 - T'(y_{i,H})\right] y_{i,L}(1+e)\frac{d\ln(w_{i,L})}{d\ln\left[1 - T'(y_{i,H})\right]} + \frac{\alpha e}{1+\alpha e}\frac{d\ln(w_{i,L})}{d\ln\left[1 - T'(y_{i,H})\right]} \\ & - \frac{\alpha}{1+\alpha e}\eta_{i,H}\left[1 - T'(y_{i,H})\right] y_{i,H}\left[(1+e)\frac{d\ln(w_{i,H})}{d\ln\left[1 - T'(y_{i,H})\right]} + e\right] - \frac{\alpha e}{1+\alpha e}.\end{aligned}$$

Using the expression for a low-skilled worker's wage response, one can rewrite a high-skilled worker's wage change as

$$\frac{d\ln(w_{i,H})}{d\ln\left[1 - T'(y_{i,H})\right]} = -\frac{\alpha e}{1+e}\frac{1 + \eta_{i,H}\left[1 - T'(y_{i,H})\right] y_{i,H}}{1 + (1-\alpha)\eta_{i,L}\left[1 - T'(y_{i,H})\right] y_{i,L} + \alpha\eta_{i,H}\left[1 - T'(y_{i,H})\right] y_{i,H}}.$$

Therefore, recalling that $\alpha < 1$, the relationship between high-skilled workers' income and the retention rate of the top tax rate is positive

$$\frac{d\ln\left(y_{i,H}\right)}{d\ln\left[1-T'\left(y_{i,H}\right)\right]}=-\alpha e\frac{1+\eta_{i,H}\left[1-T'\left(y_{i,H}\right)\right]y_{i,H}}{1+\left(1-\alpha\right)\eta_{i,L}\left[1-T'\left(y_{i,H}\right)\right]y_{i,L}+\alpha\eta_{i,H}\left[1-T'\left(y_{i,H}\right)\right]y_{i,H}}+e>0.$$

A.3 Coordinated Tax Policy

We consider a situation in which the two governments can jointly set their country-specific tax schedules to maximize world welfare. Then, the world social planner chooses $\{c_{i,\theta},l_{i,\theta},N_{i,\theta}\}_{\theta\in\Theta,i=A,B}$ to maximize

$$\sum_{i=A,B}u\left(c_{i,1},l_{i,1}\right) \tag{24}$$

subject to the high-skilled workers' incentive constraints (Equation (5)), each country's resource constraint (Equation (6)), the endogeneity of wages (Equation (1)), and the equilibrium population (Equation (3)).

Observe that, as before, the set of constraints needs to hold at a country level.[28] Then, the Lagrangian of the outer problem reads as

$$\begin{aligned}\mathcal{L}\left(\{N_{i,\theta}\}_{\theta\in\Theta,i=A,B}\right)\equiv&\sum_{i=A,B}u\left(c_{i,1},l_{i,1}\right)+\sum_{i=A,B}\sum_{\theta\in\{2,\ldots,K\}}\mu_{i,\theta}\left[u\left(c_{i,\theta},l_{i,\theta}\right)-u\left(c_{i,\theta-1},\frac{l_{i,\theta-1}w_{i,\theta-1}}{w_{i,\theta}}\right)\right]\\&+\sum_{i=A,B}\xi_i\left[F_i\left(\{N_{i,\theta}l_{i,\theta}\}_{\theta\in\Theta}\right)-\sum_{\theta\in\Theta}N_{i,\theta}c_{i,\theta}\right]+\sum_{i=A,B}\sum_{\theta\in\Theta}\lambda_{i,\theta}\left[N_{i,\theta}-\rho_i\left(\Delta_{i,\theta};\theta\right)\right],\end{aligned}$$

which yields the following first-order conditions

$$\begin{aligned}\left[c_{i,\theta}\right]:0=&\,\mathbb{1}\left[\theta=1\right]u_c\left(c_{i,\theta},l_{i,\theta}\right)-N_{i,\theta}\xi_i+\mathbb{1}\left[\theta>1\right]\mu_{i,\theta}u_c\left(c_{i,\theta},l_{i,\theta}\right)-\mathbb{1}\left[\theta<K\right]\mu_{i,\theta+1}u_c\left(c_{i,\theta},\frac{y_{i,\theta}}{w_{i,\theta+1}}\right)\\&-\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}u_c\left(c_{i,\theta},l_{i,\theta}\right)+\lambda_{j,\theta}\eta_{j,\theta}N_{j,\theta}u_c\left(c_{j,\theta},l_{j,\theta}\right)\end{aligned} \tag{25}$$

$$\begin{aligned}\left[l_{i,\theta}\right]:0=&\,\mathbb{1}\left[\theta=1\right]u_l\left(c_{i,\theta},l_{i,\theta}\right)+\mathbb{1}\left[\theta>1\right]\mu_{i,\theta}u_l\left(c_{i,\theta},l_{i,\theta}\right)-\mathbb{1}\left[\theta<K\right]\mu_{i,\theta+1}u_l\left(c_{i,\theta},\frac{y_{i,\theta}}{w_{i,\theta+1}}\right)\frac{w_{i,\theta}}{w_{i,\theta+1}}\\&+\xi_iw_{i,\theta}N_{i,\theta}-\lambda_{i,\theta}\eta_{i,\theta}N_{i,\theta}u_l\left(c_{i,\theta},l_{i,\theta}\right)+\lambda_{j,\theta}\eta_{j,\theta}N_{j,\theta}u_l\left(c_{j,\theta},l_{j,\theta}\right)\\&-\sum_{k\in\{2,\ldots,K\}}\mu_{i,k}u_l\left(c_{i,k-1},\frac{y_{i,k-1}}{w_{i,k}}\right)\frac{y_{i,k-1}}{y_{i,\theta}}\left(\gamma_{i,k-1,\theta}-\gamma_{i,k,\theta}\right),\end{aligned} \tag{26}$$

[28] Alternatively, one could consider a planner problem where the aggregate resource constraint (6) only has to hold worldwide, allowing governments to achieve cross-country redistribution by trading consumption levels. Although straightforward to consider, we disregard such incentives for the sake of comparability and due to their limited feasibility.

for $i = A, B$ and $\theta \in \Theta$. Observe that the world social planner now takes into account cross-country externalities from international migration.

As before, plug (25) into (26) and use the workers' first-order condition as well as the CES production function to get the worker's marginal tax rate

$$T_i^{co\prime}(y_{i,\theta}) = \frac{\mathbb{1}[\theta < K] \frac{\mu_{i,\theta+1}^{co}}{\xi_i^{co}} \frac{u_c\left(c_{i,\theta}, \frac{l_{i,\theta} w_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}} \left[1 + \frac{u_l\left(c_{i,\theta}, \frac{l_{i,\theta} w_{i,\theta}}{w_{i,\theta+1}}\right)}{w_{i,\theta+1} u_c\left(c_{i,\theta}, l_{i,\theta}\right)} \frac{\sigma - 1}{\sigma}\right]}{1 + \mathbb{1}[\theta < K] \frac{\mu_{i,\theta+1}^{co}}{\xi_i^{co}} \frac{u_c\left(c_{i,\theta}, \frac{l_{i,\theta} w_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}}$$
$$+ \frac{\mathbb{1}[\theta > 1] \frac{\mu_{i,\theta}^{co}}{\xi_i^{co}} \frac{y_{i,\theta-1}}{y_{i,\theta} N_{i,\theta}} \frac{1}{\sigma} \frac{u_l\left(c_{i,\theta-1}, \frac{l_{i,\theta-1} w_{i,\theta-1}}{w_{i,\theta}}\right)}{w_{i,\theta}}}{1 + \mathbb{1}[\theta < K] \frac{\mu_{i,\theta+1}^{co}}{\xi_i^{co}} \frac{u_c\left(c_{i,\theta}, \frac{l_{i,\theta} w_{i,\theta}}{w_{i,\theta+1}}\right)}{N_{i,\theta}}},$$

where $\frac{\mu_{i,\theta}^{co}}{\xi_i^{co}}$ denote the relative Lagrangian multipliers under tax coordination. Then, sum up Equations (25) over all types and plug the resulting expression for

$$\xi_i^{co} = \frac{1 - \sum_{\theta \in \Theta} \lambda_{i,\theta} \eta_{i,\theta} N_{i,\theta} + \sum_{\theta \in \Theta} \lambda_{j,\theta} \eta_{j,\theta} N_{j,\theta} \frac{u_c\left(c_{j,\theta}, l_{j,\theta}\right)}{u_c\left(c_{i,\theta}, l_{i,\theta}\right)}}{\sum_{\theta \in \Theta} N_{i,\theta} / u_c\left(c_{i,\theta}, l_{i,\theta}\right)}$$

back into (25) to solve for $\mu_{i,\theta}^{co}$.

Under cross-country symmetry, the marginal value of public funds is the same as the one of the exogenous technology planner, $\xi_i^{co} = \xi_i^{ex}$. Moreover, Equation (25) simplifies to

$$0 = \mathbb{1}[\theta = 1] u_c(c_{i,\theta}, l_{i,\theta}) - N_{i,\theta} \xi_i^{ex} + \mathbb{1}[\theta > 1] \mu_{i,\theta}^{co} u_c(c_{i,\theta}, l_{i,\theta}) - \mathbb{1}[\theta < K] \mu_{i,\theta+1}^{co} u_c\left(c_{i,\theta}, \frac{l_{i,\theta} w_{i,\theta}}{w_{i,\theta+1}}\right).$$

This implies that the other Lagrange multipliers coincide as well, $\mu_{i,\theta}^{co} = \mu_{i,\theta}^{ex}$, and the coordination solution is equivalent to the autarky allocation: $T_i^{co\prime}(y_{i,\theta}) = T_i^{ex\prime}(y_{i,\theta})$.

B Proofs for the Continuous-Type Model

B.1 Monotonicity

We order types according to their Nash equilibrium wage rates. To ensure that the type ordering is without loss of generality, we make the following assumption.

Assumption 1. *For all $\theta \in \Theta$, let $\left|p_i\left(y_{i,\theta}\right)\varepsilon_{i,\theta}^{l,(1-\tau)}\right| < 1$, $\varepsilon_{i,\theta}^{l,w}/\sigma > -1$, and suppose that $T_i'\left(y_{i,\theta}\right) \leq 1$.*

The first inequality ensures that the elasticities along the nonlinear budget line are well-defined. The second and third inequalities imply a monotone mapping between types and wages in an open economy with migration, so that the ordering of types we impose is without loss of generality. This observation is, for instance, also necessary to ensure that, in the calibration of Section 4, the ordering of types is the same for the calibrated wage distribution and the one evaluated at the optimal tax system.

For a CES production function, a worker θ's wage is given by Equation (1) in the K-type setup (Equation (8) in the continuous-type framework). As the number of types becomes very large, and thus the distance between types infinitesimal (as in the continuous-type setup of Section 3), we can write

$$\frac{\hat{w}_{i,\theta}}{w_{i,\theta}} = \frac{\hat{a}_{i,\theta}}{a_{i,\theta}} - \frac{1}{\sigma}\left[\frac{\hat{l}_{i,\theta}}{l_{i,\theta}} + \frac{\hat{N}_{i,\theta}}{N_{i,\theta}}\right].$$

Then, note that

$$\frac{\hat{l}_{i,\theta}}{l_{i,\theta}} = \varepsilon_{i,\theta}^{l,w}\frac{\hat{w}_{i,\theta}}{w_{i,\theta}} \quad \text{and} \quad \frac{\hat{N}_{i,\theta}}{N_{i,\theta}} = \eta_{i,\theta}y_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right)\frac{\hat{w}_{i,\theta}}{w_{i,\theta}} \tag{27}$$

such that

$$\frac{\hat{w}_{i,\theta}}{w_{i,\theta}}\left[1 + \frac{1}{\sigma}\left(\varepsilon_{i,\theta}^{l,w} + \eta_{i,\theta}y_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right)\right)\right] = \frac{\hat{a}_{i,\theta}}{a_{i,\theta}}.$$

Therefore, the type ordering is preserved, i.e., $\frac{\hat{w}_{i,\theta}}{w_{i,\theta}}$ and $\frac{\hat{a}_{i,\theta}}{a_{i,\theta}}$ have the same sign whenever

$$1 + \frac{1}{\sigma}\left(\varepsilon_{i,\theta}^{l,w} + \eta_{i,\theta}y_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right)\right) > 0$$

This condition is satisfied for $\varepsilon_{i,\theta}^{l,w}/\sigma > -1$ and $T_i'\left(y_{i,\theta}\right) \leq 1$ (Assumption 1). The intuition is that, according to (27), workers' labor supply and their equilibrium population closely follow their wages. Thus, when a particular type, say the low-skilled, becomes arbitrarily scarce due to emigration, their wage rate would ceteris paribus explode. However, in response, the labor supply of those still residing in the country would rise (to infinity) and there would be a large inflow of workers of this type. This aggregate labor supply adjustment would then mitigate the wage change.

Thus, the ordering of wages is the same as the ordering of productivity types after any (possibly large) tax reform. Finally, observe that, by the Spence-Mirrlees condition, $\varepsilon_{i,\theta}^{l,w} > -1$.

Hence, there is also a monotone mapping between wages and income, as $\frac{\hat{y}_{i,\theta}}{y_{i,\theta}} = \left(1 + \varepsilon_{i,\theta}^{l,w}\right) \frac{\hat{w}_{i,\theta}}{w_{i,\theta}}$.

B.2 Proof of Lemma 1

In the following, we derive the incidence of an arbitrary tax reform on endogenous labor supply, wages, and migration. Labor supply and migration responses depend on wage effects that materialize in general equilibrium (and vice versa). Thus, our first goal is to derive an expression for the general-equilibrium incidence on wages. Then, we use the resulting closed-form expressions to derive the aggregate effects of a tax change on government revenues.

As described in the main text, labor supply responds to changes in the marginal tax and the wage rate

$$\frac{\hat{l}_{i,\theta}}{l_{i,\theta}} = \frac{\hat{l}_{i,\theta}^{PE}}{l_{i,\theta}} + \varepsilon_{i,\theta}^{l,w} \frac{\hat{w}_{i,\theta}}{w_{i,\theta}} = \varepsilon_{i,\theta}^{l,(1-T')} \frac{-\hat{T}_i'\left(y_{i,\theta}\right)}{1 - T_i'\left(y_{i,\theta}\right)} + \varepsilon_{i,\theta}^{l,w} \frac{\hat{w}_{i,\theta}}{w_{i,\theta}}. \tag{28}$$

Similarly, perturbing the equilibrium population reveals a response to a change in a worker's tax payment and her wage

$$\frac{\hat{N}_{i,\theta}}{N_{i,\theta}} = \frac{\hat{N}_{i,\theta}^{PE}}{N_{i,\theta}} + \eta_{i,\theta} y_{i,\theta} \left(1 - T_i'\left(y_{i,\theta}\right)\right) \frac{\hat{w}_{i,\theta}}{w_{i,\theta}} = -\hat{T}_i\left(y_{i,\theta}\right) \eta_{i,\theta} + \eta_{i,\theta} y_{i,\theta} \left(1 - T_i'\left(y_{i,\theta}\right)\right) \frac{\hat{w}_{i,\theta}}{w_{i,\theta}}. \tag{29}$$

There are no behavioral effects on the equilibrium population (Envelope theorem). As a next step, we perturb the wage equation (8)

$$\frac{\hat{w}_{i,\theta}}{w_{i,\theta}} = -\frac{1}{\sigma}\left(\frac{\hat{l}_{i,\theta}}{l_{i,\theta}} + \frac{\hat{N}_{i,\theta}}{N_{i,\theta}}\right) + \frac{1}{\sigma} \frac{\int_{\theta \in \Theta} a_{i,\theta} \left(\frac{\hat{l}_{i,\theta}}{l_{i,\theta}} + \frac{\hat{N}_{i,\theta}}{N_{i,\theta}}\right) \left(l_{i,\theta} N_{i,\theta}\right)^{\frac{\sigma-1}{\sigma}} d\theta}{\int_{\theta \in \Theta} a_{i,\theta} \left(l_{i,\theta} N_{i,\theta}\right)^{\frac{\sigma-1}{\sigma}} d\theta}. \tag{30}$$

By combining the labor supply and population responses derived in (28) and (29), we can rewrite the equation for the wage response (30). The combined labor supply and population response is

$$\frac{\hat{l}_{i,\theta}}{l_{i,\theta}}+\frac{\hat{N}_{i,\theta}}{N_{i,\theta}}=\frac{\hat{l}^{PE}_{i,\theta}}{l_{i,\theta}}+\frac{\hat{N}^{PE}_{i,\theta}}{N_{i,\theta}}-\frac{1}{\sigma}\left(\varepsilon^{l,w}_{i,\theta}+\eta_{i,\theta}y_{i,\theta}\left(1-T'_i\left(y_{i,\theta}\right)\right)\right)\left(\frac{\hat{l}_{i,\theta}}{l_{i,\theta}}+\frac{\hat{N}_{i,\theta}}{N_{i,\theta}}\right)$$
$$+\frac{1}{\sigma}\left(\varepsilon^{l,w}_{i,\theta}+\eta_{i,\theta}y_{i,\theta}\left(1-T'_i\left(y_{i,\theta}\right)\right)\right)\frac{\int_{\theta\in\Theta}a_{i,\theta}\left(\frac{\hat{l}_{i,\theta}}{l_{i,\theta}}+\frac{\hat{N}_{i,\theta}}{N_{i,\theta}}\right)\left(l_{i,\theta}N_{i,\theta}\right)^{\frac{\sigma-1}{\sigma}}d\theta}{\int_{\theta\in\Theta}a_{i,\theta}\left(l_{i,\theta}N_{i,\theta}\right)^{\frac{\sigma-1}{\sigma}}d\theta}$$
$$=\frac{1}{1+\frac{1}{\sigma}\left(\varepsilon^{l,w}_{i,\theta}+\eta_{i,\theta}y_{i,\theta}\left(1-T'_i\left(y_{i,\theta}\right)\right)\right)}\left(\frac{\hat{l}^{PE}_{i,\theta}}{l_{i,\theta}}+\frac{\hat{N}^{PE}_{i,\theta}}{N_{i,\theta}}\right)$$
$$+\frac{\frac{1}{\sigma}\left(\varepsilon^{l,w}_{i,\theta}+\eta_{i,\theta}y_{i,\theta}\left(1-T'_i\left(y_{i,\theta}\right)\right)\right)}{1+\frac{1}{\sigma}\left(\varepsilon^{l,w}_{i,\theta}+\eta_{i,\theta}y_{i,\theta}\left(1-T'_i\left(y_{i,\theta}\right)\right)\right)}\frac{\int_{\theta\in\Theta}a_{i,\theta}\left(\frac{\hat{l}_{i,\theta}}{l_{i,\theta}}+\frac{\hat{N}_{i,\theta}}{N_{i,\theta}}\right)\left(l_{i,\theta}N_{i,\theta}\right)^{\frac{\sigma-1}{\sigma}}d\theta}{\int_{\theta\in\Theta}a_{i,\theta}\left(l_{i,\theta}N_{i,\theta}\right)^{\frac{\sigma-1}{\sigma}}d\theta}.$$

The last term can be integrated out

$$\frac{\int_{\theta\in\Theta}a_{i,\theta}\left(\frac{\hat{l}_{i,\theta}}{l_{i,\theta}}+\frac{\hat{N}_{i,\theta}}{N_{i,\theta}}\right)\left(l_{i,\theta}N_{i,\theta}\right)^{\frac{\sigma-1}{\sigma}}d\theta}{\int_{\theta\in\Theta}a_{i,\theta}\left(l_{i,\theta}N_{i,\theta}\right)^{\frac{\sigma-1}{\sigma}}d\theta}$$
$$=\frac{\int_{\theta\in\Theta}\left(\frac{dl^{PE}_{i,\theta}}{l_{i,\theta}}+\frac{dN^{PE}_{i,\theta}}{N_{i,\theta}}\right)y_{i,\theta}N_{i,\theta}/\left[1+\frac{1}{\sigma}\left(\varepsilon^{l,w}_{i,\theta}+\eta_{i,\theta}y_{i,\theta}\left(1-T'_i\left(y_{i,\theta}\right)\right)\right)\right]d\theta}{\int_{\theta\in\Theta}y_{i,\theta}N_{i,\theta}/\left[1+\frac{1}{\sigma}\left(\varepsilon^{l,w}_{i,\theta}+\eta_{i,\theta}y_{i,\theta}\left(1-T'_i\left(y_{i,\theta}\right)\right)\right)\right]d\theta}$$

and inserted back into the wage equation (30)

$$\frac{\hat{w}_{i,\theta}}{w_{i,\theta}}=-\frac{1}{\sigma}\frac{1}{1+\frac{1}{\sigma}\left(\varepsilon^{l,w}_{i,\theta}+\eta_{i,\theta}y_{i,\theta}\left(1-T'\left(y_{i,\theta}\right)\right)\right)}\left\{\frac{\hat{l}^{PE}_{i,\theta}}{l_{i,\theta}}+\frac{\hat{N}^{PE}_{i,\theta}}{N_{i,\theta}}\right.$$
$$\left.-\frac{\int_{\theta\in\Theta}\left(\frac{dl^{PE}_{i,\theta}}{l_{i,\theta}}+\frac{dN^{PE}_{i,\theta}}{N_{i,\theta}}\right)y_{i,\theta}N_{i,\theta}/\left[1+\frac{1}{\sigma}\left(\varepsilon^{l,w}_{i,\theta}+\eta_{i,\theta}y_{i,\theta}\left(1-T'_i\left(y_{i,\theta}\right)\right)\right)\right]d\theta}{\int_{\theta\in\Theta}y_{i,\theta}N_{i,\theta}/\left[1+\frac{1}{\sigma}\left(\varepsilon^{l,w}_{i,\theta}+\eta_{i,\theta}y_{i,\theta}\left(1-T'_i\left(y_{i,\theta}\right)\right)\right)\right]d\theta}\right\}. \tag{31}$$

Now, we take the Gateaux derivative of tax revenues to obtain the aggregate first-order effect

$$\hat{\mathcal{R}}_i=\int_{\theta\in\Theta}\hat{T}_i\left(y_{i,\theta}\right)N_{i,\theta}d\theta+\int_{\theta\in\Theta}\frac{\hat{l}_{i,\theta}}{l_{i,\theta}}y_{i,\theta}T'_i\left(y_{i,\theta}\right)N_{i,\theta}d\theta-\int_{\theta\in\Theta}T_i\left(y_{i,\theta}\right)\hat{T}_i\left(y_{i,\theta}\right)\eta_{i,\theta}N_{i,\theta}d\theta$$
$$+\int_{\theta\in\Theta}\frac{\hat{w}^{STW}_{i,\theta}}{w_{i,\theta}}y_{i,\theta}T'_i\left(y_{i,\theta}\right)N_{i,\theta}d\theta+\int_{\theta\in\Theta}\left(\frac{\hat{w}_{i,\theta}}{w_{i,\theta}}-\frac{\hat{w}^{STW}_{i,\theta}}{w_{i,\theta}}\right)y_{i,\theta}T'_i\left(y_{i,\theta}\right)N_{i,\theta}d\theta$$
$$+\int_{\theta\in\Theta}T_i\left(y_{i,\theta}\right)\frac{\hat{w}_{i,\theta}}{w_{i,\theta}}y_{i,\theta}\left(1-T'_i\left(y_{i,\theta}\right)\right)\eta_{i,\theta}N_{i,\theta}d\theta,$$

which is the effect decomposition in Lemma 1. Plug in the solution to the wage incidence equation (31) into the revenue differential

$$\hat{\mathcal{R}}_i = \int_{\theta\in\Theta} \left(1-\eta_{i,\theta}T_i\left(y_{i,\theta}\right)\right)\hat{T}_i\left(y_{i,\theta}\right)N_{i,\theta}d\theta + \int_{\theta\in\Theta} T_i'\left(y_{i,\theta}\right)\varepsilon_{i,\theta}^{l,(1-T')}\frac{-\hat{T}_i'\left(y_{i,\theta}\right)}{1-T_i'\left(y_{i,\theta}\right)}y_{i,\theta}N_{i,\theta}d\theta$$
$$-\frac{1}{\sigma}\int_{\theta\in\Theta}\frac{\left(1-\eta_{i,\theta}T_i\left(y_{i,\theta}\right)+\varepsilon_{i,\theta}^{l,w}\right)T_i'\left(y_{i,\theta}\right)+\eta_{i,\theta}T_i\left(y_{i,\theta}\right)}{1+\frac{1}{\sigma}\left[\varepsilon_{i,\theta}^{l,w}+y_{i,\theta}\left(1-T_i'\left(y_{i,\theta}\right)\right)\eta_{i,\theta}\right]}\left(\varepsilon_{i,\theta}^{l,(1-T')}\frac{-\hat{T}_i'\left(y_{i,\theta}\right)}{1-T_i'\left(y_{i,\theta}\right)}-\eta_{i,\theta}\hat{T}_i\left(y_{i,\theta}\right)\right)y_{i,\theta}N_{i,\theta}d\theta$$
$$+\frac{A}{\sigma}\int_{\theta\in\Theta}\frac{\varepsilon_{i,\theta}^{l,(1-T')}\frac{-\hat{T}_i'\left(y_{i,\theta}\right)}{1-T_i'\left(y_{i,\theta}\right)}-\eta_{i,\theta}\hat{T}_i\left(y_{i,\theta}\right)}{1+\frac{1}{\sigma}\left[\varepsilon_{i,\theta}^{l,w}+y_{i,\theta}\left(1-T_i'\left(y_{i,\theta}\right)\right)\eta_{i,\theta}\right]}y_{i,\theta}N_{i,\theta}d\theta,$$

where

$$A \equiv \frac{\int_{\theta\in\Theta}\left[\left(1-\eta_{i,\theta}T_i\left(y_{i,\theta}\right)+\varepsilon_{i,\theta}^{l,w}\right)T_i'\left(y_{i,\theta}\right)+\eta_{i,\theta}T_i\left(y_{i,\theta}\right)\right]y_{i,\theta}N_{i,\theta}/\left(1+\frac{1}{\sigma}\left[\varepsilon_{i,\theta}^{l,w}+y_{i,\theta}\left(1-T_i'\left(y_{i,\theta}\right)\right)\eta_{i,\theta}\right]\right)d\theta}{\int_{\theta\in\Theta}y_{i,\theta}N_{i,\theta}/\left(1+\frac{1}{\sigma}\left[\varepsilon_{i,\theta}^{l,w}+y_{i,\theta}\left(1-T_i'\left(y_{i,\theta}\right)\right)\eta_{i,\theta}\right]\right)d\theta}. \tag{32}$$

One can now use this general expression that applies to any (well-defined) tax reform of an arbitrary initial tax code to study a specific tax reform of the current tax code.

B.3 Proof of Corollary 1

Under the isoleastic disutility of labor and the CRP tax schedule defined in the main text, the labor supply elasticities are constant $\varepsilon_{i,\theta}^{l,(1-T')} = \frac{e}{1+p_i e}$ and $\varepsilon_{i,\theta}^{l,w} = (1-p_i)\,\varepsilon_{i,\theta}^{l,(1-T')}$. Moreover, under the assumption of a constant migration elasticity, it can be expressed as $\nu_i = c_{i,\theta}\eta_{i,\theta} = \frac{1-\tau_i}{1-p_i}y_{i,\theta}^{1-p_i}\eta_{i,\theta}$. In particular, observe that a one-percent increase in wages raises each type's population by a constant percentage share $y_{i,\theta}\left(1-T_i'\left(y_{i,\theta}\right)\right)\eta_{i,\theta} = (1-p_i)\,\nu_i$.

We consider the perturbation by Saez (2001) that raises tax payments above a certain income level y^*: $\hat{T}_i\left(y_{i,\theta}\right) = \frac{\mathbb{1}\left(y_{i,\theta}\geq y^*\right)}{1-\int_{\theta\leq\theta^*}N_{i,\theta}d\theta}$ and $\hat{T}_i'\left(y_{i,\theta}\right) = \frac{\delta\left(y_{i,\theta}-y^*\right)}{1-\int_{\theta\leq\theta^*}N_{i,\theta}d\theta}$ where $\delta\left(\cdot\right)$ is the Dirac delta function and θ^* is the type just earning y^*.

Plug the tax reform and the preference specifications into the wage effect on migration

$$\mathcal{WEM}_i = \int_{\theta\in\Theta} T_i\left(y_{i,\theta}\right)\frac{\hat{w}_{i,\theta}}{w_{i,\theta}}y_{i,\theta}\left(1-T_i'\left(y_{i,\theta}\right)\right)\eta_{i,\theta}N_{i,\theta}d\theta = (1-p_i)\,\nu_i\int_{\theta\in\Theta}\left(T_i\left(y_{i,\theta}\right)-y_{i,\theta}\right)\frac{\hat{w}_{i,\theta}}{w_{i,\theta}}N_{i,\theta}d\theta$$

to show that

$$\mathcal{WEM}_i = \frac{\frac{1}{\sigma}\nu_i}{1+\frac{1}{\sigma}\left(\varepsilon_i^{l,w}+(1-p_i)\,\nu_i\right)} \frac{\varepsilon_i^{l,(1-T')} y_{i,\theta^*} N_{i,\theta^*}}{1-\int_{\theta\le\theta^*} N_{i,\theta} d\theta} \frac{\int_{\theta'\in\Theta}\left(\left(\frac{y_{i,\theta^*}}{y_{i,\theta'}}\right)^{p_i}-1\right) y_{i,\theta'} N_{i,\theta'} d\theta'}{\int_{\theta\in\Theta} y_{i,\theta} N_{i,\theta} d\theta}$$
$$+\frac{\frac{1}{\sigma}\nu_i}{1+\frac{1}{\sigma}\left(\varepsilon_i^{l,w}+(1-p_i)\,\nu_i\right)}(1-p_i)\,\nu_i \int_{\theta>\theta^*} \frac{\int_{\theta'\in\Theta}\left(\left(\frac{y_{i,\theta}}{y_{i,\theta'}}\right)^{p_i}-1\right) y_{i,\theta'} N_{i,\theta'} d\theta'}{\int_{\theta\in\Theta} y_{i,\theta} N_{i,\theta} d\theta} \frac{N_{i,\theta} d\theta}{1-\int_{\theta\le\theta^*} N_{i,\theta} d\theta}.$$

Note that $\left(\frac{y_{i,\theta^*}}{y_{i,\theta'}}\right)^{p_i}-1 = \frac{T'\left(y_{i,\theta^*}\right)-T'\left(y_{i,\theta'}\right)}{1-T'\left(y_{i,\theta^*}\right)}$. Finally, change the integrals from types θ to incomes $y_{i,\theta}$ to obtain the expression in Corollary 1, where

$$t_{1,i} \equiv \frac{\frac{1}{\sigma}\nu_i e}{1+p_i e+\frac{1}{\sigma}(1-p_i)\left(e+(1+p_i e)\,\nu_i\right)}$$

and

$$t_{2,i} \equiv \nu_i \frac{\frac{1}{\sigma}(1-p_i)(1+p_i e)\,\nu_i}{1+p_i e+\frac{1}{\sigma}(1-p_i)\left(e+(1+p_i e)\,\nu_i\right)}.$$

Similarly, the migration effect on wages

$$\begin{aligned}\mathcal{MEW}_i &= \int_{\theta\in\Theta}\left(1+\varepsilon_{i,\theta}^{l,w}\right)\left(\frac{\hat{w}_{i,\theta}}{w_{i,\theta}}-\frac{\hat{w}_{i,\theta}^{STW}}{w_{i,\theta}}\right) y_{i,\theta} T_i'\left(y_{i,\theta}\right) N_{i,\theta} d\theta \\ &= \left(1+\varepsilon_i^{l,w}\right)\int_{\theta\in\Theta}\left(\frac{\hat{w}_{i,\theta}}{w_{i,\theta}}-\frac{\hat{w}_{i,\theta}^{STW}}{w_{i,\theta}}\right) y_{i,\theta}\left(T_i'\left(y_{i,\theta}\right)-1\right) N_{i,\theta} d\theta\end{aligned}$$

simplifies to

$$\mathcal{MEW}_i = -\frac{\left(1+\varepsilon_i^{l,w}\right)\frac{1}{\sigma}\varepsilon_i^{l,w}/\varepsilon_i^{l,(1-T')}}{1+\frac{1}{\sigma}\varepsilon^{l,w}} \frac{\frac{1}{\sigma}\nu_i}{1+\frac{1}{\sigma}\left(\varepsilon_i^{l,w}+(1-p_i)\,\nu_i\right)}$$
$$\cdot\frac{\varepsilon_i^{l,(1-T')} y_{i,\theta^*} N_{i,\theta^*}}{1-\int_{\theta\le\theta^*} N_{i,\theta} d\theta} \frac{\int_{\theta'\in\Theta}\left(\left(\frac{y_{i,\theta^*}}{y_{i,\theta'}}\right)^{p_i}-1\right) y_{i,\theta'} N_{i,\theta'} d\theta'}{\int_{\theta\in\Theta} y_{i,\theta} N_{i,\theta} d\theta}$$
$$+\frac{\left(1+\varepsilon_i^{l,w}\right)\frac{1}{\sigma}(1-p_i)\,\nu_i}{1+\frac{1}{\sigma}\left(\varepsilon_i^{l,w}+(1-p_i)\,\nu_i\right)} \int_{\theta>\theta^*} \frac{\int_{\theta'\in\Theta}\left(\left(\frac{y_{i,\theta}}{y_{i,\theta'}}\right)^{p_i}-1\right) y_{i,\theta'} N_{i,\theta'} d\theta'}{\int_{\theta\in\Theta} y_{i,\theta} N_{i,\theta} d\theta} \frac{N_{i,\theta} d\theta}{1-\int_{\theta\le\theta^*} N_{i,\theta} d\theta},$$

from which the expression in the corollary directly follows with

$$t_{3,i} \equiv t_{1,i}\cdot\frac{\frac{1}{\sigma}(1-p_i)(1+e)}{1+p_i e+\frac{1}{\sigma}(1-p_i)\,e}$$

and

$$t_{4,i} \equiv t_{2,i} \cdot \frac{1+e}{(1+p_i e)\,\nu_i}.$$

B.4 Proof of Proposition 2

The proof of Proposition 2 follows the same steps as in Sachs et al. (2020). Again, we consider an elementary tax reform (Saez (2001)): $\hat{T}_i(y_{i,\theta}) = \frac{\mathbb{1}(y_{i,\theta}\geq y^*)}{1-\int_{\theta\leq\theta^*} N_{i,\theta}d\theta}$ and $\hat{T}_i'(y_{i,\theta}) = \frac{\delta(y_{i,\theta}-y^*)}{1-\int_{\theta\leq\theta^*} N_{i,\theta}d\theta}$. For this tax reform, the aggregate incidence reads as

$$\begin{aligned}
\hat{\mathcal{R}}_i = & \int_{\theta>\theta^*} \left(1-\eta_{i,\theta}T_i(y_{i,\theta})\right) N_{i,\theta}\frac{d\theta}{1-\int_{\theta\leq\theta^*} N_{i,\theta}d\theta} - \frac{y_{i,\theta^*}N_{i,\theta^*}}{1-\int_{\theta\leq\theta^*} N_{i,\theta}d\theta}\varepsilon_{i,\theta^*}^{l,(1-T')}\frac{T_i'(y_{i,\theta^*})}{1-T_i'(y_{i,\theta^*})} \\
& + \frac{1}{\sigma}\frac{\left(1-\eta_{i,\theta^*}T_i(y_{i,\theta^*})+\varepsilon_{i,\theta^*}^{l,w}\right)T_i'(y_{i,\theta^*})+\eta_{i,\theta^*}T_i(y_{i,\theta^*})}{1+\frac{1}{\sigma}\left[\varepsilon_{i,\theta^*}^{l,w}+y_{i,\theta^*}\left(1-T_i'(y_{i,\theta^*})\right)\eta_{i,\theta^*}\right]}\frac{y_{i,\theta^*}N_{i,\theta^*}}{1-\int_{\theta\leq\theta^*} N_{i,\theta}d\theta}\varepsilon_{i,\theta^*}^{l,(1-T')}\frac{1}{1-T_i'(y_{i,\theta^*})} \\
& + \frac{1}{\sigma}\int_{\theta>\theta^*}\frac{\left(1-\eta_{i,\theta}T_i(y_{i,\theta})+\varepsilon_{i,\theta}^{l,w}\right)T_i'(y_{i,\theta})+\eta_{i,\theta}T_i(y_{i,\theta})}{1+\frac{1}{\sigma}\left[\varepsilon_{i,\theta}^{l,w}+y_{i,\theta}\left(1-T_i'(y_{i,\theta})\right)\eta_{i,\theta}\right]}\eta_{i,\theta}y_{i,\theta}N_{i,\theta}\frac{d\theta}{1-\int_{\theta\leq\theta^*} N_{i,\theta}d\theta} \\
& - \frac{A}{\sigma}\frac{1}{1+\frac{1}{\sigma}\left[\varepsilon_{i,\theta^*}^{l,w}+y_{i,\theta^*}\left(1-T_i'(y_{i,\theta^*})\right)\eta_{i,\theta^*}\right]}\frac{y_{i,\theta^*}N_{i,\theta^*}}{1-\int_{\theta\leq\theta^*} N_{i,\theta}d\theta}\varepsilon_{i,\theta^*}^{l,(1-T')}\frac{1}{1-T_i'(y_{i,\theta^*})} \\
& - \frac{A}{\sigma}\int_{\theta>\theta^*}\frac{1}{1+\frac{1}{\sigma}\left[\varepsilon_{i,\theta}^{l,w}+y_{i,\theta}\left(1-T_i'(y_{i,\theta})\right)\eta_{i,\theta}\right]}\eta_{i,\theta}y_{i,\theta}N_{i,\theta}\frac{d\theta}{1-\int_{\theta\leq\theta^*} N_{i,\theta}d\theta},
\end{aligned}$$

where the term A was defined above in (32). Collect terms to get

$$\begin{aligned}
\hat{\mathcal{R}}_i = & \int_{\theta>\theta^*}\left(1-\eta_{i,\theta}T_i(y_{i,\theta})+\frac{1}{\sigma}\frac{\left(1-\eta_{i,\theta}T_i(y_{i,\theta})\right)\left(T_i'(y_{i,\theta})-1\right)+\varepsilon_{i,\theta}^{l,w}T_i'(y_{i,\theta})-(A-1)}{1+\frac{1}{\sigma}\left[\varepsilon_{i,\theta}^{l,w}+y_{i,\theta}\left(1-T_i'(y_{i,\theta})\right)\eta_{i,\theta}\right]}\eta_{i,\theta}y_{i,\theta}\right)\frac{N_{i,\theta}d\theta}{1-\int_{\theta\leq\theta^*} N_{i,\theta}d\theta} \\
& - \frac{T_i'(y_{i,\theta^*})\left(1+\frac{1}{\sigma}y_{i,\theta^*}\left(1-T_i'(y_{i,\theta^*})\right)\eta_{i,\theta^*}\right)+\frac{1}{\sigma}\left(1-T_i'(y_{i,\theta^*})\right)\left(1-\eta_{i,\theta^*}T_i(y_{i,\theta^*})\right)+\frac{1}{\sigma}(A-1)}{\left(1+\frac{1}{\sigma}\left[\varepsilon_{i,\theta^*}^{l,w}+y_{i,\theta^*}\left(1-T_i'(y_{i,\theta^*})\right)\eta_{i,\theta^*}\right]\right)\left(1-T_i'(y_{i,\theta^*})\right)}\frac{\varepsilon_{i,\theta^*}^{l,(1-T')}y_{i,\theta^*}N_{i,\theta^*}}{1-\int_{\theta\leq\theta^*} N_{i,\theta}d\theta}.
\end{aligned}$$

In the optimum, it must be that there are no first-order effects, $\hat{\mathcal{R}}_i = 0$, for any tax reform (also the one considered above). We now show that this is the case when $A = 1$. Changing from types to incomes and imposing $A=1$, the optimal tax scheme is given by

$$\begin{aligned}\frac{T_i'(y^*)}{1-T_i'(y^*)}=&\frac{1+\frac{1}{\sigma}\left[\varepsilon_{i,y^*}^{l,w}+y^*\left(1-T_i'(y^*)\right)\eta_{i,y^*}\right]}{\varepsilon_{i,y^*}^{l,(1-T')}}\frac{1-\int_{y'\leq y^*}N_{i,y'}dy'}{y^*N_{i,y^*}}\\&\times\int_{y>y^*}\left[1-\eta_{i,y}T_i(y)+\frac{1}{\sigma}\eta_{i,y}y\frac{\left(1-\eta_{i,y}T_i(y)\right)\left(T_i'(y)-1\right)+\varepsilon_{i,y}^{l,w}T_i'(y)}{1+\frac{1}{\sigma}\left[\varepsilon_{i,y}^{l,w}+y\left(1-T_i'(y)\right)\eta_{i,y}\right]}\right]\frac{N_ydy}{1-\int_{y'\leq y^*}N_{y'}dy}\\&-\frac{1-y^*\eta_{i,y^*}\left(T_i(y^*)/y^*-T_i'(y^*)\right)}{\sigma}\end{aligned}$$

To show that $A=1$, we consider an alternative reform: $\hat{T}_i(y_{i,\theta})=-K_i(y^*)\left(1-T_i'(y_{i,\theta})\right)y_{i,\theta}$ and $\hat{T}_i'(y_{i,\theta})=-K_i(y^*)\left(1-T_i'(y_{i,\theta})-y_{i,\theta}T_i''(y_{i,\theta})\right)$. Accordingly, plug

$$\eta_{i,\theta}\hat{T}_i(y_{i,\theta})=-K_i(y^*)\left(1-T_i'(y_{i,\theta})\right)y_{i,\theta}\eta_{i,\theta}$$

and

$$\varepsilon_{i,\theta}^{l,(1-T')}\frac{-\hat{T}_i'(y_{i,\theta})}{1-T_i'(y_{i,\theta})}=K_i(y^*)\,\varepsilon_{i,\theta}^{l,w}$$

into $\hat{\mathcal{R}}_i=0$ and collect terms

$$\begin{aligned}0=&\int_{\theta\in\Theta}\frac{\left[\left(1-\eta_{i,\theta}T_i(y_{i,\theta})+\varepsilon_{i,\theta}^{l,w}\right)T_i'(y_{i,\theta})+\eta_{i,\theta}T_i(y_{i,\theta})\right]}{1+\frac{1}{\sigma}\left[\varepsilon_{i,\theta}^{l,w}+y_{i,\theta}\left(1-T_i'(y_{i,\theta})\right)\eta_{i,\theta}\right]}y_{i,\theta}N_{i,\theta}d\theta\\&+(A-1)\int_{\theta\in\Theta}y_{i,\theta}N_{i,\theta}d\theta-A\int_{\theta\in\Theta}\frac{1}{1+\frac{1}{\sigma}\left[\varepsilon_{i,\theta}^{l,w}+y_{i,\theta}\left(1-T_i'(y_{i,\theta})\right)\eta_{i,\theta}\right]}y_{i,\theta}N_{i,\theta}d\theta\\&=(A-1)\int_{\theta\in\Theta}y_{i,\theta}N_{i,\theta}d\theta,\end{aligned}$$

which concludes the proof.

B.5 Alternative Welfare Criteria

Departing from a Rawlsian social planner, the individual incidence on labor supply, wages, and migration, as studied in Section B.2, remains unchanged. Under a social welfare criterion that adds up residents' welfare (see Saez and Stantcheva (2016)), each government solves

$$\max_{\left\{T_i\left(y_{i,\theta}\right)\right\}_{\theta\in\Theta}}\mathcal{G}_i\quad\text{subject to}\quad\mathcal{R}_i\equiv\int_{\theta\in\Theta}T_i(y_{i,\theta})\,N_{i,\theta}d\theta\geq0,$$

where

$$\mathcal{G}_i \equiv \int_{\theta\in\Theta} \mathit{\Gamma}_i\left(\mathcal{U}_{i,\theta}\right) N_{i,\theta} d\theta$$

is the aggregate social welfare in country i and $\mathit{\Gamma}_i : \mathbb{R} \to \mathbb{R}_+$ defines an increasing and concave generalized social welfare function. Again, denote $\xi_i = \int_{\theta\in\Theta} \mathit{\Gamma}_i'\left(\mathcal{U}_{i,\theta}\right) u_{i,\theta}' N_{i,\theta} d\theta$ as the marginal value of public funds.

To obtain the aggregate first-order effect, we now add up welfare and tax revenue effects

$$\begin{aligned}
\hat{\mathcal{G}}_i/\xi_i + \hat{\mathcal{R}}_i = & \int_{\theta\in\Theta} \hat{T}_i\left(y_{i,\theta}\right) N_{i,\theta} d\theta + \int_{\theta\in\Theta} \frac{\hat{l}_{i,\theta}}{l_{i,\theta}} y_{i,\theta} T_i'\left(y_{i,\theta}\right) N_{i,\theta} d\theta - \int_{\theta\in\Theta} T_i\left(y_{i,\theta}\right) \hat{T}_i\left(y_{i,\theta}\right) \eta_{i,\theta} N_{i,\theta} d\theta \\
& + \int_{\theta\in\Theta} \frac{\hat{w}_{i,\theta}}{w_{i,\theta}} y_{i,\theta} T_i'\left(y_{i,\theta}\right) N_{i,\theta} d\theta + \int_{\theta\in\Theta} T_i\left(y_{i,\theta}\right) \frac{\hat{w}_{i,\theta}}{w_{i,\theta}} y_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right) \eta_{i,\theta} N_{i,\theta} d\theta \\
& - \int_{\theta\in\Theta} \left(\mathit{\Gamma}_i'\left(\mathcal{U}_{i,\theta}\right) u_{i,\theta}'/\xi_i + \mathit{\Gamma}_i\left(\mathcal{U}_{i,\theta}\right)/\xi_i \eta_{i,\theta}\right) \hat{T}_i\left(y_{i,\theta}\right) N_{i,\theta} d\theta \\
& + \int_{\theta\in\Theta} \left(\mathit{\Gamma}_i'\left(\mathcal{U}_{i,\theta}\right) u_{i,\theta}'/\xi_i + \mathit{\Gamma}_i\left(\mathcal{U}_{i,\theta}\right)/\xi_i \eta_{i,\theta}\right) \frac{\hat{w}_{i,\theta}}{w_{i,\theta}} y_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right) \eta_{i,\theta} N_{i,\theta} d\theta,
\end{aligned}$$

which is similar to the effect decomposition in (13). As before, plug in the solution to the wage incidence equation (12) into this differential

$$\begin{aligned}
\hat{\mathcal{G}}_i/\xi_i + \hat{\mathcal{R}}_i = & \int_{\theta\in\Theta} \left(1 - \mathit{\Gamma}_i'\left(\mathcal{U}_{i,\theta}\right) \mathcal{U}_{i,\theta}'/\xi_i - \eta_{i,\theta}\left(T_i\left(y_{i,\theta}\right) + \mathit{\Gamma}_i\left(\mathcal{U}_{i,\theta}\right)/\xi_i\right)\right) \hat{T}_i\left(y_{i,\theta}\right) N_{i,\theta} d\theta \\
& + \int_{\theta\in\Theta} T_i'\left(y_{i,\theta}\right) \varepsilon_{i,\theta}^{l,(1-T')} \frac{-\hat{T}_i{}'\left(y_{i,\theta}\right)}{1 - T_i'\left(y_{i,\theta}\right)} y_{i,\theta} N_{i,\theta} d\theta \\
& - \frac{1}{\sigma} \int_{\theta\in\Theta} \mathit{\Omega}_{i,\theta} \frac{\varepsilon_{i,\theta}^{l,(1-T')} \frac{-\hat{T}_i'\left(y_{i,\theta}\right)}{1-T_i'\left(y_{i,\theta}\right)} - \eta_{i,\theta} \hat{T}_i\left(y_{i,\theta}\right)}{1 + \frac{1}{\sigma}\left[\varepsilon_{i,\theta}^{l,w} + y_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right) \eta_{i,\theta}\right]} y_{i,\theta} N_{i,\theta} d\theta \\
& + \frac{A}{\sigma} \int_{\theta\in\Theta} \frac{\varepsilon_{i,\theta}^{l,(1-T')} \frac{-\hat{T}_i'\left(y_{i,\theta}\right)}{1-T_i'\left(y_{i,\theta}\right)} - \eta_{i,\theta} \hat{T}_i\left(y_{i,\theta}\right)}{1 + \frac{1}{\sigma}\left[\varepsilon_{i,\theta}^{l,w} + y_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right) \eta_{i,\theta}\right]} y_{i,\theta} N_{i,\theta} d\theta,
\end{aligned}$$

where

$$A \equiv \frac{\int_{\theta\in\Theta} \mathit{\Omega}_{i,\theta} y_{i,\theta} N_{i,\theta} / \left(1 + \frac{1}{\sigma}\left[\varepsilon_{i,\theta}^{l,w} + y_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right) \eta_{i,\theta}\right]\right) d\theta}{\int_{\theta\in\Theta} y_{i,\theta} N_{i,\theta} / \left(1 + \frac{1}{\sigma}\left[\varepsilon_{i,\theta}^{l,w} + y_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right) \eta_{i,\theta}\right]\right) d\theta}$$

and

$$\mathit{\Omega}_{i,\theta} \equiv \left(1 - \eta_{i,\theta}\left(T_i\left(y_{i,\theta}\right) + \mathit{\Gamma}_i\left(\mathcal{U}_{i,\theta}\right)/\xi_i\right) + \varepsilon_{i,\theta}^{l,w}\right) T_i'\left(y_{i,\theta}\right) + \mathit{\Gamma}_i'\left(\mathcal{U}_{i,\theta}\right) u_{i,\theta}'/\xi_i + \eta_{i,\theta}\left(T_i\left(y_{i,\theta}\right) + \mathit{\Gamma}_i\left(\mathcal{U}_{i,\theta}\right)/\xi_i\right).$$

To solve for an ABC-formula for the optimal income tax scheme, we set the aggregate effects of an elementary tax reform equal to zero, $\hat{\mathcal{G}}_i/\xi_i + \hat{\mathcal{R}}_i = 0$, similar to before. Then, we get

$$\begin{aligned}
\hat{\mathcal{G}}_i/\xi_i + \hat{\mathcal{R}}_i = 0 = & \int_{\theta>\theta^*} \left(1 - \eta_{i,\theta} T_i\left(y_{i,\theta}\right)\right) N_{i,\theta} \frac{d\theta}{1 - \int_{\theta\leq\theta^*} N_{i,\theta} d\theta} - \frac{y_{i,\theta^*} N_{i,\theta^*}}{1 - \int_{\theta\leq\theta^*} N_{i,\theta} d\theta} \varepsilon_{i,\theta^*}^{l,(1-T')} \frac{T_i'\left(y_{i,\theta^*}\right)}{1 - T_i'\left(y_{i,\theta^*}\right)} \\
& + \frac{1}{\sigma} \frac{\Omega_{i,\theta^*}}{1 + \frac{1}{\sigma}\left[\varepsilon_{i,\theta^*}^{l,w} + y_{i,\theta^*}\left(1 - T_i'\left(y_{i,\theta^*}\right)\right)\eta_{i,\theta^*}\right]} \frac{y_{i,\theta^*} N_{i,\theta^*}}{1 - \int_{\theta\leq\theta^*} N_{i,\theta} d\theta} \varepsilon_{i,\theta^*}^{l,(1-T')} \frac{1}{1 - T_i'\left(y_{i,\theta^*}\right)} \\
& + \frac{1}{\sigma} \int_{\theta>\theta^*} \frac{\Omega_{i,\theta}}{1 + \frac{1}{\sigma}\left[\varepsilon_{i,\theta}^{l,w} + y_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right)\eta_{i,\theta}\right]} \eta_{i,\theta} y_{i,\theta} N_{i,\theta} \frac{d\theta}{1 - \int_{\theta\leq\theta^*} N_{i,\theta} d\theta} \\
& - \frac{A}{\sigma} \frac{1}{1 + \frac{1}{\sigma}\left[\varepsilon_{i,\theta^*}^{l,w} + y_{i,\theta^*}\left(1 - T_i'\left(y_{i,\theta^*}\right)\right)\eta_{i,\theta^*}\right]} \frac{y_{i,\theta^*} N_{i,\theta^*}}{1 - \int_{\theta\leq\theta^*} N_{i,\theta} d\theta} \varepsilon_{i,\theta^*}^{l,(1-T')} \frac{1}{1 - T_i'\left(y_{i,\theta^*}\right)} \\
& - \frac{A}{\sigma} \int_{\theta>\theta^*} \frac{1}{1 + \frac{1}{\sigma}\left[\varepsilon_{i,\theta}^{l,w} + y_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right)\eta_{i,\theta}\right]} \eta_{i,\theta} y_{i,\theta} N_{i,\theta} \frac{d\theta}{1 - \int_{\theta\leq\theta^*} N_{i,\theta} d\theta}.
\end{aligned}$$

or, equivalently,

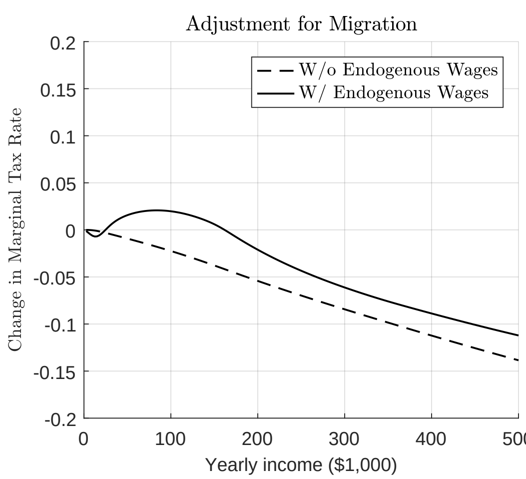
$$\begin{aligned}
0 = & \int_{\theta>\theta^*} \Big(1 - \Gamma_i'\left(\mathcal{U}_{i,\theta}\right) u_{i,\theta}'/\xi_i - \eta_{i,\theta}\left(T_i\left(y_{i,\theta}\right) + \Gamma_i\left(\mathcal{U}_{i,\theta}\right)/\xi_i\right) \\
& + \frac{1}{\sigma} \frac{\left(1 - \eta_{i,\theta}\left(T_i\left(y_{i,\theta}\right) + \Gamma_i\left(\mathcal{U}_{i,\theta}\right)/\xi_i\right)\right)\left(T_i'\left(y_{i,\theta}\right) - 1\right) + \varepsilon_{i,\theta}^{l,w} T_i'\left(y_{i,\theta}\right) - (A-1)}{1 + \frac{1}{\sigma}\left[\varepsilon_{i,\theta}^{l,w} + y_{i,\theta}\left(1 - T_i'\left(y_{i,\theta}\right)\right)\eta_{i,\theta}\right]} \eta_{i,\theta} y_{i,\theta}\Bigg) \frac{N_{i,\theta} d\theta}{1 - \int_{\theta\leq\theta^*} N_{i,\theta} d\theta} \\
& - \left(T_i'\left(y_{i,\theta^*}\right)\left(1 + \frac{1}{\sigma} y_{i,\theta^*}\left(1 - T_i'\left(y_{i,\theta^*}\right)\right)\eta_{i,\theta^*}\right) + \frac{1}{\sigma}\left(1 - T_i'\left(y_{i,\theta^*}\right)\right)\right. \\
& \left. \cdot \left(1 - \Gamma_i'\left(\mathcal{U}_{i,\theta^*}\right) u_{i,\theta^*}'/\xi_i - \eta_{i,\theta^*}\left(T_i\left(y_{i,\theta^*}\right) + \Gamma_i\left(\mathcal{U}_{i,\theta^*}\right)/\xi_i\right)\right) + \frac{1}{\sigma}(A-1)\right) \\
& \times \frac{1}{\left(1 + \frac{1}{\sigma}\left[\varepsilon_{i,\theta^*}^{l,w} + y_{i,\theta^*}\left(1 - T_i'\left(y_{i,\theta^*}\right)\right)\eta_{i,\theta^*}\right]\right)\left(1 - T_i'\left(y_{i,\theta^*}\right)\right)} \frac{\varepsilon_{i,\theta^*}^{l,(1-T')} y_{i,\theta^*} N_{i,\theta^*}}{1 - \int_{\theta\leq\theta^*} N_{i,\theta} d\theta}.
\end{aligned}$$

Imposing that $A = 1$ and changing from types to incomes, we get a formula for the optimal tax schedule

$$
\begin{aligned}
\frac{T_i'(y^*)}{1-T_i'(y^*)} = & \frac{1+\frac{1}{\sigma}\left[\varepsilon_{i,y^*}^{l,w}+y^*\left(1-T_i'(y^*)\right)\eta_{i,y^*}\right]}{\varepsilon_{i,y^*}^{l,(1-T')}} \frac{1-\int_{y'\leq y^*} N_{i,y'}dy'}{y^* N_{i,y^*}} \\
& \times \int_{y>y^*} \left[1-\Gamma_i'\left(\mathcal{U}_{i,y}\right)u_{i,y}'/\xi_i - \eta_{i,y}\left(T_i(y)+\Gamma_i\left(\mathcal{U}_{i,y}\right)/\xi_i\right)\right. \\
& \left. +\frac{1}{\sigma}\eta_{i,y} y \frac{\left(1-\Gamma_i'\left(\mathcal{U}_{i,y}\right)u_{i,y}'/\xi_i - \eta_{i,y}\left(T_i(y)+\Gamma_i\left(\mathcal{U}_{i,y}\right)/\xi_i\right)\right)\left(T_i'(y)-1\right)+\varepsilon_{i,y}^{l,w}T_i'(y)}{1+\frac{1}{\sigma}\left[\varepsilon_{i,y}^{l,w}+y\left(1-T_i'(y)\right)\eta_{i,y}\right]}\right] \\
& - \frac{1-\Gamma_i'\left(\mathcal{U}_{i,y^*}\right)u_{i,y^*}'/\xi_i - y^*\eta_{i,y^*}\left(T_i(y^*)/y^* - T_i'(y^*)+\Gamma_i\left(\mathcal{U}_{i,y^*}\right)/\xi_i\right)}{\sigma}
\end{aligned}
$$

To show that $A=1$, consider again an alternative reform $\hat{T}_i(y_{i,\theta}) = -K_i(y^*)\left(1-T_i'(y_{i,\theta})\right)y_{i,\theta}$ and $\hat{T}_i'(y_{i,\theta}) = -K_i(y^*)\left(1-T_i'(y_{i,\theta})-y_{i,\theta}T_i''(y_{i,\theta})\right)$, plug it into $\hat{\mathcal{G}}_i/\xi_i+\hat{\mathcal{R}}_i=0$, and use the definition of A.

C Further Simulation Results



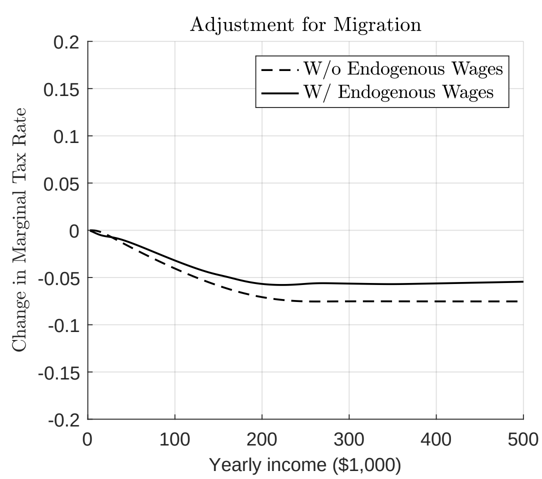


Figure 5: Adjustment in Optimal Taxation for Migration in PE and GE; Left Panel: Constant Migration Semi-Elasticity; Right Panel: Decreasing Migration Semi-Elasticity

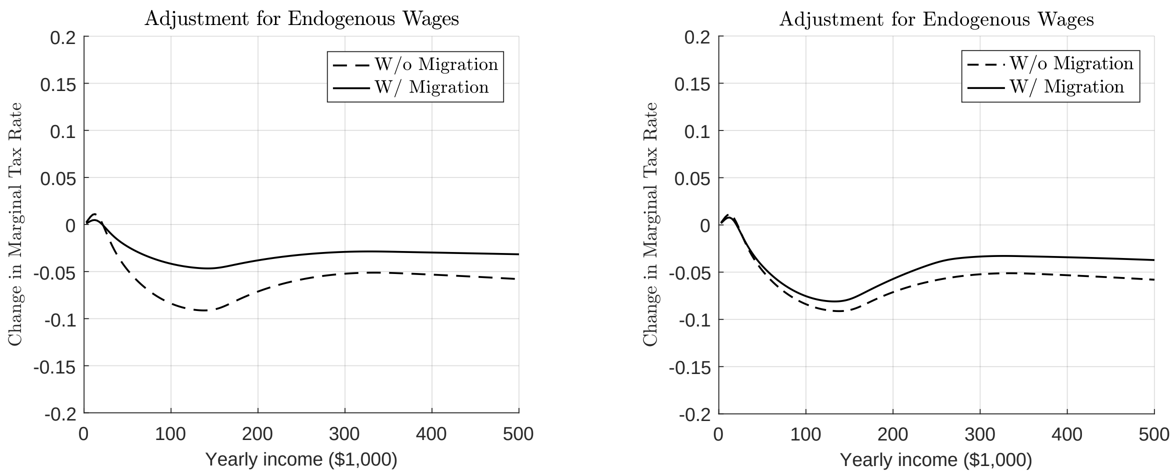


Figure 6: Adjustment in Optimal Taxation for Endogenous Wages W/ and W/o Migration; Left Panel: Constant Migration Semi-Elasticity; Right Panel: Decreasing Migration Semi-Elasticity